Graduate Texts in Mathematics 261

GTM

Erhan Çınlar

Probability and Stochastics

Springer

Graduate Texts in Mathematics 261

For other titles in this series, go to
http://www.springer.com/series/136

Erhan Çınlar

Probability and Stochastics

Springer

Erhan Çınlar
Princeton University
328 Sherrerd Hall
Princeton, NJ 08544
USA
ecinlar@princeton.edu

ISSN 0072-5285
ISBN 978-0-387-87858-4 e-ISBN 978-0-387-87859-1
DOI 10.1007/978-0-387-87859-1
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011921929

Mathematics Subject Classification (2010): 60

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This is an introduction to the modern theory of probability and stochastic processes. The aim is to enable the student to have access to the many excellent research monographs in the literature. It might be regarded as an updated version of the textbooks by Breiman, Chung, and Neveu, just to name three.

The book is based on the lecture notes for a two-semester course which I have offered for many years. The course is fairly popular and attracts graduate students in engineering, economics, physics, and mathematics, and a few overachieving undergraduates. Most of the students had familiarity with elementary probability, but it was safer to introduce each concept carefully and in a uniform style.

As Martin Barlow put it once, mathematics attracts us because the need to memorize is minimal. So, only the more fundamental facts are labeled as theorems; they are worth memorizing. Most other results are put as propositions, comments, or exercises. Also put as exercises are results that can be understood only by doing the tedious work necessary. I believe in the Chinese proverb: I hear, I forget; I see, I remember; I do, I know.

I have been considerate: I do not assume that the reader will go through the book line by line from the beginning to the end. Some things are recalled or re-introduced when they are needed. In each chapter or section, the essential material is put first, technical material is put toward the end. Subheadings are used to introduce the subjects and results; the reader should have a quick overview by flipping the pages and reading the headings.

The style and coverage is geared toward the theory of stochastic processes, but with some attention to the applications. The reader will find many instances where the gist of the problem is introduced in practical, everyday language, and then is made precise in mathematical form. Conversely, many a theoretical point is re-stated in heuristic terms in order to develop the intuition and to provide some experience in stochastic modeling.

The first four chapters are on the classical probability theory: random variables, expectations, conditional expectations, independence, and the classical limit theorems. This is more or less the minimum required in a course at graduate level probability. There follow chapters on martingales, Poisson random measures, Lévy processes, Brownian motion, and Markov processes.

The first chapter is a review of measure and integration. The treatment is in tune with the modern literature on probability and stochastic processes. The second chapter introduces probability spaces as special measure spaces, but with an entirely different emotional effect; sigma-algebras are equated to bodies of information, and measurability to determinability by the given information. Chapter III is on convergence; it is routinely classical; it goes through the definitions of different modes of convergence, their connections to each other, and the classical limit theorems. Chapter IV is on conditional expectations as estimates given some information, as projection operators, and as Radon-Nikodym derivatives. Also in this chapter is the construction of probability spaces using conditional probabilities as the initial data.

Martingales are introduced in Chapter V in the form initiated by P.-A. Meyer, except that the treatment of continuous martingales seems to contain an improvement, achieved through the introduction of a "Doob martingale", a stopped martingale that is uniformly integrable. Also in this chapter are two great theorems: martingale characterization of Brownian motion due to Lévy and the martingale characterization of Poisson process due to Watanabe.

Poisson random measures are developed in Chapter VI with some care. The treatment is from the point of view of their uses in the study of point processes, discontinuous martingales, Markov processes with jumps, and, especially, of Lévy processes. As the modern theory pays more attention to processes with jumps, this chapter should fulfill an important need. Various uses of them occur in the remaining three chapters.

Chapter VII is on Lévy processes. They are treated as additive processes just as Lévy and Itô thought of them. Itô-Lévy decomposition is presented fully, by following Itô's method, thus laying bare the roles of Brownian motion and Poisson random measures in the structure of Lévy processes and, with a little extra thought, the structure of most Markov processes. Subordination of processes and the hitting times of subordinators are given extra attention.

Chapter VIII on Brownian motion is mostly on the standard material: hitting times, the maximum process, local times, and excursions. Poisson random measures are used to clarify the structure of local times and Itô's characterization of excursions. Also, Bessel processes and some other Markov processes related to Brownian motion are introduced; they help explain the recurrence properties of Brownian motion, and they become examples for the Markov processes to be introduced in the last chapter.

Chapter IX is the last, on Markov processes. Itô diffusions and jump-diffusions are introduced via stochastic integral equations, thus displaying the process as an integral path in a field of Lévy processes. For such processes, we derive the classical relationships between martingales, generators, resolvents, and transition functions, thus introducing the analytic theory of them. Then we re-introduce Markov processes in the modern setting and explain, for Hunt processes, the meaning and implications of the strong Markov property and quasi-left-continuity.

Over the years, I have acquired indebtedness to many students for their enthusiastic search for errors in the manuscript. In particular, Semih Sezer and Yury Polyanskiy were helpful with corrections and improved proofs. The manuscript was formatted by Emmanuel Sharef in his junior year, and Willie Wong typed the first six chapters during his junior and senior years. Siu-Tang Leung typed the seventh chapter, free of charge, out of sheer kindness. Evan Papageorgiou prepared the figures on Brownian motion and managed the latex files for me. Finally, Springer has shown much patience as I missed deadline after deadline, and the staff there did an excellent job with the production. Many thanks to all.

Frequently Used Notation

$\mathbb{N} = \{0, 1, \ldots\}$, $\overline{\mathbb{N}} = \{0, 1, \ldots, +\infty\}$, $\mathbb{N}^* = \{1, 2, \ldots\}$.

$\mathbb{R} = (-\infty, +\infty)$, $\overline{\mathbb{R}} = [-\infty, +\infty]$, $\mathbb{R}_+ = [0, \infty)$, $\overline{\mathbb{R}}_+ = [0, +\infty]$.

$(a,\ b)$ is the open interval with endpoints a and b; the closed version is $[a,\ b]$; the left-open right-closed version is $(a,\ b]$.

$\exp\ x = e^x$, $\exp_\ x = e^{-x}$, Leb is the Lebesgue measure.

$\mathbb{R}^d$ is the d-dimensional Euclidean space, for x and y in it,

$$x \cdot y = x_1 y_1 + \cdots + x_d y_d, \quad |x| = \sqrt{x \cdot x} \quad .$$

$(E, \mathcal{E})$ denotes a measurable space, $\mathcal{E}$ is also the set of all $\mathcal{E}$-measurable functions from E into $\overline{\mathbb{R}}$, and $\mathcal{E}_+$ is the set of positive functions in $\mathcal{E}$.

$1_A(x) = \delta_x(A) = I(x, A)$ is equal to 1 if $x \in A$ and to 0 otherwise.

$\mathcal{B}_E$ is the Borel σ-algebra on E when E is topological.

$C(E \mapsto F)$ is the set of all continuous functions from E into F.

$\mathcal{C}_K^2 = C_K^2(\mathbb{R}^d \mapsto \mathbb{R})$ is the set of twice continuously differentiable functions, from $\mathbb{R}^d$ into $\mathbb{R}$, with compact support.

$\mathbb{E}(X|\mathcal{G})$ is the conditional expectation of X given the σ-algebra $\mathcal{G}$.

$\mathbb{E}_t X = \mathbb{E}(X|\mathcal{F}_t)$ when the filtration $(\mathcal{F}_t)$ is held fixed.

Contents

Preface **v**

Frequently Used Notation **ix**

I Measure and Integration **1**
- 1 Measurable Spaces 1
- 2 Measurable Functions 6
- 3 Measures 14
- 4 Integration 19
- 5 Transforms and Indefinite Integrals 29
- 6 Kernels and Product Spaces 37

II Probability Spaces **49**
- 1 Probability Spaces and Random Variables 50
- 2 Expectations 57
- 3 L^p-spaces and Uniform Integrability 70
- 4 Information and Determinability 75
- 5 Independence 82

III Convergence **93**
- 1 Convergence of Real Sequences 93
- 2 Almost Sure Convergence 97
- 3 Convergence in Probability 101
- 4 Convergence in L^p 105
- 5 Weak Convergence 109
- 6 Laws of Large Numbers 118
- 7 Convergence of Series 124
- 8 Central Limits 127

IV Conditioning **139**
- 1 Conditional Expectations 139
- 2 Conditional Probabilities and Distributions 149
- 3 Conditional Independence 158
- 4 Construction of Probability Spaces 160
- 5 Special Constructions 166

V Martingales and Stochastics **171**
1 Filtrations and Stopping Times 171
2 Martingales 181
3 Martingale Transformations and Maxima 190
4 Martingale Convergence 199
5 Martingales in Continuous Time 213
6 Martingale Characterizations for Wiener and Poisson 225
7 Standard Filtrations and Modifications of Martingales 234

VI Poisson Random Measures **243**
1 Random Measures 243
2 Poisson Random Measures 248
3 Transformations 263
4 Additive Random Measures and Lévy Processes 277
5 Poisson Processes 290
6 Poisson Integrals and Self-exciting Processes 298

VII Lévy Processes **313**
1 Introduction 313
2 Stable Processes 329
3 Lévy Processes on Standard Settings 340
4 Characterizations for Wiener and Poisson 349
5 Itô-Lévy Decomposition 354
6 Subordination 360
7 Increasing Lévy Processes 368

VIII Brownian Motion **379**
1 Introduction 379
2 Hitting Times and Recurrence Times 389
3 Hitting Times and Running Maximum 396
4 Wiener and its Maximum 399
5 Zeros, Local Times 408
6 Excursions 413
7 Path Properties 426
8 Existence 437

IX Markov Processes **443**
1 Markov Property 444
2 Itô Diffusions 455
3 Jump-Diffusions 473
4 Markov Systems 498
5 Hunt Processes 505
6 Potentials and Excessive Functions 518
7 Appendix: Stochastic Integration 525

Notes and Comments **533**

Bibliography **541**

Index **551**

Chapter I

Measure and Integration

This chapter is devoted to the basic notions of measurable spaces, measure, and integration. The coverage is limited to what probability theory requires as the entrance fee from its students. The presentation is in the form and style attuned to the modern treatments of probability theory and stochastic processes.

1 Measurable Spaces

Let E be a set. We use the usual notations for operations on subsets of E:

1.1 $$A \cup B, \quad A \cap B, \quad A \setminus B$$

denote, respectively, the union of A and B, the intersection of A and B, and the complement of B in A. In particular, $E \setminus B$ is called simply the complement of B and is also denoted by B^c. We write $A \subset B$ or $B \supset A$ to mean that A is a subset of B, that is, A is contained in B, or equivalently, B contains A. Note that $A = B$ if and only if $A \subset B$ and $A \supset B$. For an arbitrary collection $\{A_i : i \in I\}$ of subsets of E, we write

1.2 $$\bigcup_{i \in I} A_i, \qquad \bigcap_{i \in I} A_i$$

for the union and intersection, respectively, of all the sets A_i, $i \in I$.

The empty set is denoted by $\emptyset$. Sets A and B are said to be *disjoint* if $A \cap B = \emptyset$. A collection of sets is said to be *disjointed* if its every element is disjoint from every other. A countable disjointed collection of sets whose union is A is called a *partition* of A.

A collection $\mathcal{C}$ of subsets of E is said to be *closed under* intersections if $A \cap B$ belongs to $\mathcal{C}$ whenever A and B belong to $\mathcal{C}$. Of course, then, the

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_1,

intersection of every non-empty finite collection of sets in $\mathcal{C}$ is in $\mathcal{C}$. If the intersection of every countable collection of sets in $\mathcal{C}$ is in $\mathcal{C}$, then we say that $\mathcal{C}$ is closed under countable intersections. The notions of being closed under complements, unions, and countable unions, etc. are defined similarly.

Sigma-algebras

A non-empty collection $\mathcal{E}$ of subsets of E is called an *algebra* on E provided that it be closed under finite unions and complements. It is called a *σ-algebra* on E if it is closed under complements and countable unions, that is, if

1.3 a) $A \in \mathcal{E} \quad \Rightarrow \quad E \setminus A \in \mathcal{E}$,
b) $A_1, A_2, \ldots \in \mathcal{E} \quad \Rightarrow \quad \bigcup_n A_n \in \mathcal{E}$.

Since the intersection of a collection of sets is the complement of the union of the complements of those sets, a σ-algebra is also closed under countable intersections.

Every σ-algebra on E includes E and $\emptyset$ at least. Indeed, $\mathcal{E} = \{\emptyset, E\}$ is the simplest σ-algebra on E; it is called the *trivial σ-algebra*. The largest is the collection of all subsets of E, usually denoted by 2^E; it is called the *discrete σ-algebra* on E.

The intersection of an arbitrary (countable or uncountable) family of σ-algebras on E is again a σ-algebra on E. Given an arbitrary collection $\mathcal{C}$ of subsets of E, consider all the σ-algebras that contain $\mathcal{C}$ (there is at least one such σ-algebra, namely 2^E); take the intersection of all those σ-algebras; the result is the smallest σ-algebra that contains $\mathcal{C}$; it is called the σ-algebra *generated* by $\mathcal{C}$ and is denoted by $\sigma\mathcal{C}$.

If E is a topological space, then the σ-algebra generated by the collection of all open subsets of E is called the *Borel σ-algebra* on E; it is denoted by$\mathcal{B}_E$ or $\mathcal{B}(E)$; its elements are called *Borel sets*.

p-systems and d-systems

A collection $\mathcal{C}$ of subsets of E is called a p-system if it is closed under intersections; here, p is for product, the latter being an alternative term for intersection, and next, d is for Dynkin who introduced these systems into probability. A collection $\mathcal{D}$ of subsets of E is called a d-system on E if

1.4 a) $E \in \mathcal{D}$,
b) $A, B \in \mathcal{D}$ and $A \supset B \quad \Rightarrow A \setminus B \in \mathcal{D}$,
c) $(A_n) \subset \mathcal{D}$ and $A_n \nearrow A \quad \Rightarrow A \in \mathcal{D}$.

In the last line, we wrote $(A_n) \subset \mathcal{D}$ to mean that (A_n) is a sequence of elements of $\mathcal{D}$ and we wrote $A_n \nearrow A$ to mean that the sequence is increasing with limit A in the following sense:

1.5 $$A_1 \subset A_2 \subset \ldots, \quad \cup_n A_n = A.$$

It is obvious that a σ-algebra is both a p-system and a d-system, and the converse will be shown next. Thus, p-systems and d-systems are primitive structures whose superpositions yield σ-algebras.

1.6 PROPOSITION. *A collection of subsets of E is a σ-algebra if and only if it is both a p-system and a d-system on E.*

Proof. Necessity is obvious. To show the sufficiency, let $\mathcal{E}$ be a collection of subsets of E that is both a p-system and a d-system. First, $\mathcal{E}$ is closed under complements: $A \in \mathcal{E} \Rightarrow E \setminus A \in \mathcal{E}$, since $E \in \mathcal{E}$ and $A \subset E$ and $\mathcal{E}$ is a d-system. Second, it is closed under unions: $A, B \in \mathcal{E} \Rightarrow A \cup B \in \mathcal{E}$, because $A \cup B = (A^c \cap B^c)^c$ and $\mathcal{E}$ is closed under complements (as shown) and under intersections by the hypothesis that it is a p-system. Finally, this closure extends to countable unions: if $(A_n) \subset \mathcal{E}$, then $B_1 = A_1$ and $B_2 = A_1 \cup A_2$ and so on belong to $\mathcal{E}$ by the preceding step, and $B_n \nearrow \bigcup_n A_n$, which together imply that $\bigcup_n A_n \in \mathcal{E}$ since $\mathcal{E}$ is a d-system by hypothesis. □

The lemma next is in preparation for the main theorem of this section. Its proof is left as an exercise in checking the conditions 1.4 one by one.

1.7 LEMMA. *Let $\mathcal{D}$ be a d-system on E. Fix D in $\mathcal{D}$ and let*

$$\hat{\mathcal{D}} = \{A \in \mathcal{D} : \ A \cap D \in \mathcal{D}\}$$

Then, $\hat{\mathcal{D}}$ is again a d-system.

Monotone class theorem

This is a very useful tool for showing that certain collections are σ-algebras. We give it in the form found most useful in probability theory.

1.8 THEOREM. *If a d-system contains a p-system, then it contains also the σ-algebra generated by that p-system.*

Proof. Let $\mathcal{C}$ be a p-system. Let $\mathcal{D}$ be the smallest d-system on E that contains $\mathcal{C}$, that is, $\mathcal{D}$ is the intersection of all d-systems containing $\mathcal{C}$. The claim is that $\mathcal{D} \supset \sigma\mathcal{C}$. To show it, since $\sigma\mathcal{C}$ is the smallest σ-algebra containing $\mathcal{C}$, it is sufficient to show that $\mathcal{D}$ is a σ-algebra. In view of Proposition 1.6, it is thus enough to show that the d-system $\mathcal{D}$ is also a p-system.

To that end, fix B in $\mathcal{C}$ and let

$$\mathcal{D}_1 = \{A \in \mathcal{D} : \ A \cap B \in \mathcal{D}\}.$$

Since $\mathcal{C}$ is contained in $\mathcal{D}$, the set B is in $\mathcal{D}$; and Lemma 1.7 implies that $\mathcal{D}_1$ is a d-system. It also contains $\mathcal{C}$: if $A \in \mathcal{C}$ then $A \cap B \in \mathcal{C}$ since B is in $\mathcal{C}$ and $\mathcal{C}$ is a p-system. Hence, $\mathcal{D}_1$ must contain the smallest d-system containing $\mathcal{C}$, that is, $\mathcal{D}_1 \supset \mathcal{D}$. In other words, $A \cap B \in \mathcal{D}$ for every A in $\mathcal{D}$ and B in $\mathcal{C}$.

Consequently, for fixed A in $\mathcal{D}$, the collection

$$\mathcal{D}_2 = \{B \in \mathcal{D} : \ A \cap B \in \mathcal{D}\}$$

contains $\mathcal{C}$. By Lemma 1.7, $\mathcal{D}_2$ is a d-system. Thus, $\mathcal{D}_2$ must contain $\mathcal{D}$. In other words, $A \cap B \in \mathcal{D}$ whenever A and B are in $\mathcal{D}$, that is, $\mathcal{D}$ is a p-system. □

Measurable spaces

A *measurable space* is a pair $(E, \mathcal{E})$ where E is a set and $\mathcal{E}$ is a σ-algebra on E. Then, the elements of $\mathcal{E}$ are called *measurable sets*. When E is topological and $\mathcal{E} = \mathcal{B}_E$, the Borel σ-algebra on E, then measurable sets are also called *Borel sets*.

Products of measurable spaces

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. For $A \subset E$ and $B \subset F$, we write $A \times B$ for the set of all pairs (x, y) with x in A and y in B; it is called the *product* of A and B. If $A \in \mathcal{E}$ and $B \in \mathcal{F}$, then $A \times B$ is said to be a *measurable rectangle*. We let $\mathcal{E} \otimes \mathcal{F}$ denote the σ-algebra on $E \times F$ generated by the collection of all measurable rectangles; it is called the *product σ-algebra*. The measurable space $(E \times F, \mathcal{E} \otimes \mathcal{F})$ is called the product of $(E, \mathcal{E})$ and $(F, \mathcal{F})$, and the notation $(E, \mathcal{E}) \times (F, \mathcal{F})$ is used as well.

Exercises

1.9 *Partition generated σ-algebras.*

a) Let $\mathcal{C} = \{A, B, C\}$ be a partition of E. List the elements of $\sigma\mathcal{C}$.

b) Let $\mathcal{C}$ be a (countable) partition of E. Show that every element of $\sigma\mathcal{C}$ is a countable union of elements taken from $\mathcal{C}$. Hint: Let $\mathcal{E}$ be the collection of all sets that are countable unions of elements taken from $\mathcal{C}$. Show that $\mathcal{E}$ is a σ-algebra, and argue that $\mathcal{E} = \sigma\mathcal{C}$.

c) Let $E = \mathbb{R}$, the set of all real numbers. Let $\mathcal{C}$ be the collection of all singleton subsets of $\mathbb{R}$, that is, each element of $\mathcal{C}$ is a set that consists of exactly one point in $\mathbb{R}$. Show that every element of $\sigma\mathcal{C}$ is either a countable set or the complement of a countable set. Incidentally, $\sigma\mathcal{C}$ is much smaller than $\mathcal{B}(\mathbb{R})$; for instance, the interval $(0, 1)$ belongs to the latter but not to the former.

1.10 *Comparisons.* Let $\mathcal{C}$ and $\mathcal{D}$ be two collections of subsets of E. Show the following:

a) If $\mathcal{C} \subset \mathcal{D}$ then $\sigma\mathcal{C} \subset \sigma\mathcal{D}$

b) If $\mathcal{C} \subset \sigma\mathcal{D}$ then $\sigma\mathcal{C} \subset \sigma\mathcal{D}$

c) If $\mathcal{C} \subset \sigma\mathcal{D}$ and $\mathcal{D} \subset \sigma\mathcal{C}$, then $\sigma\mathcal{C} = \sigma\mathcal{D}$

d) If $\mathcal{C} \subset \mathcal{D} \subset \sigma\mathcal{C}$, then $\sigma\mathcal{C} = \sigma\mathcal{D}$

1.11 *Borel σ-algebra on* $\mathbb{R}$. Every open subset of $\mathbb{R} = (-\infty, +\infty)$, the real line, is a countable union of open intervals. Use this fact to show that $\mathcal{B}_{\mathbb{R}}$ is generated by the collection of all open intervals.

1.12 *Continuation.* Show that every interval of $\mathbb{R}$ is a Borel set. In particular, $(-\infty, x)$, $(-\infty, x]$, $(x, y]$, $[x, y]$ are all Borel sets. For each x, the singleton $\{x\}$ is a Borel set.

1.13 *Continuation.* Show that $\mathcal{B}_{\mathbb{R}}$ is also generated by any one of the following (and many others):

a) The collection of all intervals of the form $(-\infty, x]$.
b) The collection of all intervals of the form $(x, y]$.
c) The collection of all intervals of the form $[x, y]$.
d) The collection of all intervals of the form (x, ∞).

Moreover, in each case, x and y can be limited to be rationals.

1.14 *Lemma 1.7.* Prove.

1.15 *Trace spaces.* Let $(E, \mathcal{E})$ be a measurable space. Fix $D \subset E$ and let

$$\mathcal{D} = \mathcal{E} \cap D = \{A \cap D : \ A \in \mathcal{E}\}.$$

Show that $\mathcal{D}$ is a σ-algebra on D. It is called the *trace* of $\mathcal{E}$ on D, and $(D, \mathcal{D})$ is called the *trace* of $(E, \mathcal{E})$ on D.

1.16 *Single point extensions.* Let $(E, \mathcal{E})$ be a measurable space, and let Δ be an extra point, not in E. Let $\bar{E} = E \cup \{\Delta\}$. Show that

$$\bar{\mathcal{E}} = \mathcal{E} \cup \{A \cup \{\Delta\} : \ A \in \mathcal{E}\}$$

is a σ-algebra on $\bar{E}$; it is the σ-algebra on $\bar{E}$ generated by $\mathcal{E}$.

1.17 *Product spaces.* Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Show that the product σ-algebra $\mathcal{E} \otimes \mathcal{F}$ is also the σ-algebra generated by $\hat{\mathcal{E}} \cup \hat{\mathcal{F}}$, where

$$\hat{\mathcal{E}} = \{A \times F : \ A \in \mathcal{E}\}, \quad \hat{\mathcal{F}} = \{E \times B : \ B \in \mathcal{F}\}.$$

1.18 *Unions of σ-algebras.* Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be σ-algebras on the same set E. Their union is not a σ-algebra, except in some special cases. The σ-algebra generated by $\mathcal{E}_1 \cup \mathcal{E}_2$ is denoted by $\mathcal{E}_1 \vee \mathcal{E}_2$. More generally, if $\mathcal{E}_i$ is a σ-algebra on E for each i in some (countable or uncountable) index set I, then

$$\mathcal{E}_I = \bigvee_{i \in I} \mathcal{E}_i$$

denotes the σ-algebra generated by $\bigcup_{i \in I} \mathcal{E}_i$ (a similar notation for intersection is superfluous, since $\bigcap_{i \in I} \mathcal{E}_i$ is always a σ-algebra). Let $\mathcal{C}$ be the collection of all sets A having the form

$$A = \bigcap_{i \in J} A_i$$

for some finite subset J of I and sets A_i in $\mathcal{E}_i$, $i \in J$. Show that $\mathcal{C}$ contains all $\mathcal{E}_i$ and therefore $\bigcup_I \mathcal{E}_i$. Thus, $\mathcal{C}$ generates the σ-algebra $\mathcal{E}_I$. Show that $\mathcal{C}$ is a p-system.

2 Measurable Functions

Let E and F be sets. A *mapping* or *function* f from E into F is a rule that assigns an element $f(x)$ of F to each x in E, and then we write $f: E \mapsto F$ to indicate it. If $f(x)$ is an element of F for each x in E, we also write $f: x \mapsto f(x)$ to name the mapping involved; for example, $f: x \mapsto x^2 + 5$ is the function f from $\mathbb{R}$ into $\mathbb{R}_+$ satisfying $f(x) = x^2 + 5$. Given a mapping $f: E \mapsto F$ and a subset B of F, the *inverse image* of B under f is

2.1 $$f^{-1}B = \{x \in E: \ f(x) \in B\}.$$

We leave the proof of the next lemma as an exercise in ordinary logic.

2.2 Lemma. *Let f be a mapping from E into F. Then,*

$$f^{-1}\emptyset = \emptyset, \quad f^{-1}F = E, \quad f^{-1}(B \setminus C) = (f^{-1}B) \setminus (f^{-1}C),$$

$$f^{-1}\bigcup_i B_i = \bigcup_i f^{-1}B_i, \quad f^{-1}\bigcap_i B_i = \bigcap_i f^{-1}B_i$$

for all subsets B and C of F and arbitrary collections $\{B_i: \ i \in I\}$ of subsets of F.

Measurable functions

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A mapping $f: E \mapsto F$ is said to be *measurable* relative to $\mathcal{E}$ and $\mathcal{F}$ if $f^{-1}B \in \mathcal{E}$ for every B in $\mathcal{F}$. The following reduces the checks involved.

2.3 Proposition. *In order for $f: E \mapsto F$ to be measurable relative to $\mathcal{E}$ and $\mathcal{F}$, it is necessary and sufficient that, for some collection $\mathcal{F}_0$ that generates $\mathcal{F}$, we have $f^{-1}B \in \mathcal{E}$ for every B in $\mathcal{F}_0$.*

Proof. Necessity is trivial. To prove the sufficiency, let $\mathcal{F}_0$ be a collection of subsets of F such that $\sigma\mathcal{F}_0 = \mathcal{F}$, and suppose that $f^{-1}B \in \mathcal{E}$ for every B in $\mathcal{F}_0$. We need to show that

$$\mathcal{F}_1 = \{B \in \mathcal{F}: \ f^{-1}B \in \mathcal{E}\}$$

contains $\mathcal{F}$ and thus is equal to $\mathcal{F}$. Since $\mathcal{F}_1 \supset \mathcal{F}_0$ by assumption, once we show that $\mathcal{F}_1$ is a σ-algebra, we will have $\mathcal{F}_1 = \sigma\mathcal{F}_1 \supset \sigma\mathcal{F}_0 = \mathcal{F}$ as needed. But checking that $\mathcal{F}_1$ is a σ-algebra is straightforward using Lemma 2.2. □

Composition of functions

Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, and $(G, \mathcal{G})$ be measurable spaces. Let f be a mapping from E into F, and g a mapping from F into G. The *composition* of f and g is the mapping $g \circ f$ from E into G defined by

2.4 $$g \circ f(x) = g(f(x)), \quad x \in E.$$

The next proposition will be recalled by the phrase "measurable functions of measurable functions are measurable".

2.5 PROPOSITION. *If f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and g relative to $\mathcal{F}$ and $\mathcal{G}$, then $g \circ f$ is measurable relative to $\mathcal{E}$ and $\mathcal{G}$.*

Proof. Let f and g be measurable. For C in $\mathcal{G}$, observe that $(g \circ f)^{-1}C = f^{-1}(g^{-1}C)$. Now, $g^{-1}C \in \mathcal{F}$ by the measurability of g and, hence, $f^{-1}(g^{-1}C) \in \mathcal{E}$ by the measurability of f. So, $g \circ f$ is measurable. □

Numerical functions

Let $(E, \mathcal{E})$ be a measurable space. Recall that $\mathbb{R} = (-\infty, +\infty)$, $\bar{\mathbb{R}} = [-\infty, +\infty]$, $\mathbb{R}_+ = [0, +\infty)$, $\bar{\mathbb{R}}_+ = [0, +\infty]$. A *numerical function* on E is a mapping from E into $\bar{\mathbb{R}}$ or some subset of $\bar{\mathbb{R}}$. If all its values are in $\mathbb{R}$, it is said to be *real-valued.* If all its values are in $\bar{\mathbb{R}}_+$, it is said to be *positive.*

A numerical function on E is said to be $\mathcal{E}$*-measurable* if it is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\bar{\mathbb{R}})$, the latter denoting the Borel σ-algebra on $\bar{\mathbb{R}}$ as usual. If E is topological and $\mathcal{E} = \mathcal{B}(E)$, then $\mathcal{E}$-measurable functions are called *Borel functions.*

The following proposition is a corollary of Proposition 2.3 using the fact that $\mathcal{B}(\bar{\mathbb{R}})$ is generated by the collection of intervals $[-\infty, r]$ with r in $\mathbb{R}$. No proof seems needed.

2.6 PROPOSITION. *A mapping $f : E \mapsto \bar{\mathbb{R}}$ is $\mathcal{E}$-measurable if and only if, for every r in $\mathbb{R}$, $f^{-1}[-\infty, r] \in \mathcal{E}$.*

2.7 REMARKS. a) The proposition remains true if $[-\infty, r]$ is replaced by $[-\infty, r)$ or by $[r, \infty]$ or by $(r, \infty]$, because the intervals $[-\infty, r)$ with r in $\mathbb{R}$ generate $\mathcal{B}(\bar{\mathbb{R}})$ and similarly for the other two forms.

b) In the particular case $f : E \mapsto F$, where F is a countable subset of $\bar{\mathbb{R}}$, the mapping f is $\mathcal{E}$-measurable if and only if $f^{-1}\{a\} = \{x \in E : f(x) = a\}$ is in $\mathcal{E}$ for every a in F.

Positive and negative parts of a function

For a and b in $\bar{\mathbb{R}}$ we write $a \vee b$ for the maximum of a and b, and $a \wedge b$ for the minimum. The notation extends to numerical functions naturally: for instance, $f \vee g$ is the function whose value at x is $f(x) \vee g(x)$. Let $(E, \mathcal{E})$ be a measurable space. Let f be a numerical function on E. Then,

2.8 $$f^+ = f \vee 0, \qquad f^- = -(f \wedge 0)$$

are both positive functions and $f = f^+ - f^-$. The function f^+ is called the *positive part* of f, and f^- the *negative part.*

2.9 PROPOSITION. *The function f is $\mathcal{E}$-measurable if and only if both f^+ and f^- are.*

Proof is left as an exercise. The decomposition $f = f^+ - f^-$ enables us to obtain many results for arbitrary f from the corresponding results for positive functions.

Indicators and simple functions

Let $A \subset E$. Its indicator, denoted by 1_A, is the function defined by

2.10 $$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

We write simply 1 for 1_E. Obviously, 1_A is $\mathcal{E}$-measurable if and only if $A \in \mathcal{E}$.

A function f on E is said to be *simple* if it has the form

2.11 $$f = \sum_1^n a_i 1_{A_i}$$

for some n in $\mathbb{N}^* = \{1, 2, \ldots\}$, real numbers $a_1, \ldots, a_n$, and measurable sets $A_1, \ldots, A_n$ (belonging to the σ-algebra $\mathcal{E}$). It is clear that, then, there exist m in $\mathbb{N}^*$ and distinct real numbers $b_1, \ldots, b_m$ and a measurable partition $\{B_1, \ldots, B_m\}$ of E such that $f = \sum_1^m b_i 1_{B_i}$; this latter representation is called the *canonical form* of the simple function f.

It is immediate from Proposition 2.6 (or Remark 2.7b) applied to the canonical form that every simple function is $\mathcal{E}$-measurable. Conversely, if f is $\mathcal{E}$-measurable, takes only finitely many values, and all those values are real numbers, then f is a simple function. In particular, every constant is a simple function. Finally, if f and g are simple, then so are

2.12 $$f + g, \qquad f - g, \qquad fg, \qquad f/g, \qquad f \vee g, \qquad f \wedge g,$$

except that in the case of f/g one should make sure that g is nowhere zero.

Limits of sequences of functions

Let (f_n) be a sequence of numerical functions on E. The functions

2.13 $$\inf f_n, \qquad \sup f_n, \qquad \liminf f_n, \qquad \limsup f_n$$

are defined on E pointwise: for instance, the first is the function whose value at x is the infimum of the sequence of numbers $f_n(x)$. In general, limit inferior is dominated by the limit superior. If the two are equal, that is, if

2.14 $$\liminf f_n = \limsup f_n = f,$$

say, then the sequence (f_n) is said to have a pointwise limit f and we write $f = \lim f_n$ or $f_n \to f$ to express it.

If (f_n) is increasing, that is, if $f_1 \le f_2 \le \ldots$, then $\lim f_n$ exists and is equal to $\sup f_n$. We shall write $f_n \nearrow f$ to mean that (f_n) is increasing and has limit f. Similarly, $f_n \searrow f$ means that (f_n) is decreasing and has limit f.

The following shows that the class of measurable functions is closed under limits.

2.15 THEOREM. *Let (f_n) be a sequence of $\mathcal{E}$-measurable functions. Then, each one of the four functions in 2.13 is $\mathcal{E}$-measurable. Moreover, if it exists, $\lim f_n$ is $\mathcal{E}$-measurable.*

Proof. We start by showing that $f = \sup f_n$ is $\mathcal{E}$-measurable. For every x in E and r in $\mathbb{R}$, we note that $f(x) \le r$ if and only if $f_n(x) \le r$ for all n. Thus, for each r in $\mathbb{R}$,

$$f^{-1}[-\infty, r] = \{x: \ f(x) \le r\} = \bigcap_n \{x: \ f_n(x) \le r\} = \bigcap_n f_n^{-1}[-\infty, r].$$

The rightmost member belongs to $\mathcal{E}$: for each n, the set $f_n^{-1}[-\infty, r] \in \mathcal{E}$ by the $\mathcal{E}$-measurability of f_n, and $\mathcal{E}$ is closed under countable intersections. So, by Proposition 2.6, $f = \sup f_n$ is $\mathcal{E}$-measurable.

Measurability of $\inf f_n$ follows from the preceding step upon observing that $\inf f_n = -\sup(-f_n)$. It is now obvious that

$$\liminf f_n = \sup_m \inf_{n \ge m} f_n, \qquad \limsup f_n = \inf_m \sup_{n \ge m} f_n$$

are $\mathcal{E}$-measurable. If these two are equal, the common limit is the definition of $\lim f_n$, which is $\mathcal{E}$-measurable. □

Approximation of measurable functions

We start by approximating the identity function on $\bar{\mathbb{R}}_+$ by an increasing sequence of simple functions of a specific form (dyadic functions). We leave the proof of the next lemma as an exercise; drawing d_n for $n = 1, 2, 3$ should do.

2.16 LEMMA. *For each n in $\mathbb{N}^*$, let*

$$d_n(r) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)}(r) + n 1_{[n,\infty]}(r), \qquad r \in \bar{\mathbb{R}}_+.$$

Then, each d_n is an increasing right-continuous simple function on $\bar{\mathbb{R}}_+$, and $d_n(r)$ increases to r for each r in $\bar{\mathbb{R}}_+$ as $n \to \infty$.

The following theorem is important: it reduces many a computation about measurable functions to a computation about simple functions followed by limit taking.

2.17 Theorem. *A positive function on E is $\mathcal{E}$-measurable if and only if it is the limit of an increasing sequence of positive simple functions.*

Proof. Sufficiency is immediate from Theorem 2.15. To show the necessity part, let $f : E \mapsto \bar{\mathbb{R}}_+$ be $\mathcal{E}$-measurable. We are to show that there is a sequence (f_n) of positive simple functions increasing to f. To that end, let (d_n) be as in the preceding lemma and put $f_n = d_n \circ f$. Then, for each n, the function f_n is $\mathcal{E}$-measurable, since it is a measurable function of a measurable function. Also, it is positive and takes only finitely many values, because d_n is so. Thus, each f_n is positive and simple. Moreover, since $d_n(r)$ increases to r for each r in $\bar{\mathbb{R}}_+$ as $n \to \infty$, we have that $f_n(x) = d_n(f(x))$ increases to $f(x)$ for each x in E as $n \to \infty$.

Monotone classes of functions

Let $\mathcal{M}$ be a collection of numerical functions on E. We write $\mathcal{M}_+$ for the subcollection consisting of positive functions in $\mathcal{M}$, and $\mathcal{M}_b$ for the subcollection of bounded functions in $\mathcal{M}$.

The collection $\mathcal{M}$ is called a *monotone class* provided that it includes the constant function 1, and $\mathcal{M}_b$ is a linear space over $\mathbb{R}$, and $\mathcal{M}_+$ is closed under increasing limits; more explicitly, $\mathcal{M}$ is a monotone class if

2.18
a) $1 \in \mathcal{M}$,
b) $f, g \in \mathcal{M}_b$ and $a, b \in \mathbb{R} \;\Rightarrow\; af + bg \in \mathcal{M}$,
c) $(f_n) \subset \mathcal{M}_+,\ f_n \nearrow f \;\Rightarrow\; f \in \mathcal{M}$.

The next theorem is used often to show that a certain property holds for all $\mathcal{E}$-measurable functions. It is a version of Theorem 1.8, it is called the monotone class theorem for functions.

2.19 Theorem. *Let $\mathcal{M}$ be a monotone class of functions on E. Suppose, for some p-system $\mathcal{C}$ generating $\mathcal{E}$, that $1_A \in \mathcal{M}$ for every A in $\mathcal{C}$. Then, $\mathcal{M}$ includes all positive $\mathcal{E}$-measurable functions and all bounded $\mathcal{E}$-measurable functions.*

Proof. We start by showing that $1_A \in \mathcal{M}$ for every A in $\mathcal{E}$. To this end, let

$$\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{M}\}.$$

Using the conditions 2.18, it is easy to check that $\mathcal{D}$ is a d-system. Since $\mathcal{D} \supset \mathcal{C}$ by assumption, and since $\mathcal{C}$ is a p-system that generates $\mathcal{E}$, we must have $\mathcal{D} \supset \mathcal{E}$ by the monotone class theorem 1.8. So, $1_A \in \mathcal{M}$ for every A in $\mathcal{E}$.

Therefore, in view of the property 2.18b, $\mathcal{M}$ includes all simple functions.

Let f be a positive $\mathcal{E}$-measurable function. By Theorem 2.17, there exists a sequence of positive simple functions f_n increasing to f. Since each f_n is in $\mathcal{M}_+$ by the preceding step, the property 2.18c implies that $f \in \mathcal{M}$.

Finally, let f be a bounded $\mathcal{E}$-measurable function. Then f^+ and f^- are in $\mathcal{M}$ by the preceding step and are bounded obviously. Thus, by 2.18b, we conclude that $f = f^+ - f^- \in \mathcal{M}$. □

Standard measurable spaces

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Let f be a bijection from E onto F, and let $\hat{f}$ denote its functional inverse ($\hat{f}(y) = x$ if and only if $f(x) = y$). Then, f is said to be an *isomorphism* of $(E, \mathcal{E})$ and $(F, \mathcal{F})$ if f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ and $\hat{f}$ is measurable relative to $\mathcal{F}$ and $\mathcal{E}$. The measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$ are said to be *isomorphic* if there exists an isomorphism between them.

A measurable space $(E, \mathcal{E})$ is said to be *standard* if it is isomorphic to $(F, \mathcal{B}_F)$ for some Borel subset F of $\mathbb{R}$.

The class of standard spaces is surprisingly large and includes almost all the spaces we shall encounter. Here are some examples: The spaces $\mathbb{R}$, $\mathbb{R}^d$, $\mathbb{R}^\infty$ together with their respective Borel σ-algebras are standard measurable spaces. If E is a complete separable metric space, then $(E, \mathcal{B}_E)$ is standard. If E is a Polish space, that is, if E is a topological space metrizable by a metric for which it is complete and separable, then $(E, \mathcal{B}_E)$ is standard. If E is a separable Banach space, or more particularly, a separable Hilbert space, then $(E, \mathcal{B}_E)$ is standard. Further examples will appear later.

Clearly, $[0, 1]$ and its Borel σ-algebra form a standard measurable space; so do $\{1, 2, \ldots, n\}$ and its discrete σ-algebra; so do $\mathbb{N} = \{0, 1, \ldots\}$ and its discrete σ-algebra. Every standard measurable space is isomorphic to one of these three (this is a deep result).

Notation

We shall use $\mathcal{E}$ both for the σ-algebra and for the collection of all the numerical functions that are measurable relative to it. Recall that, for an arbitrary collection $\mathcal{M}$ of numerical functions, we write $\mathcal{M}_+$ for the subcollection of positive functions in $\mathcal{M}$, and $\mathcal{M}_b$ for the subcollection of bounded ones in $\mathcal{M}$. Thus, for instance, $\mathcal{E}_+$ is the collection of all $\mathcal{E}$-measurable positive functions.

A related notation is $\mathcal{E}/\mathcal{F}$ which is used for the class of all functions $f : E \mapsto F$ that are measurable relative to $\mathcal{E}$ and $\mathcal{F}$. The notation $\mathcal{E}/\mathcal{F}$ is simplified to $\mathcal{E}$ when $F = \bar{\mathbb{R}}$ and $\mathcal{F} = \mathcal{B}(\mathbb{R})$.

Exercises and complements

2.20 *σ-algebra generated by a function.* Let E be a set and $(F, \mathcal{F})$ a measurable space. For $f : E \mapsto F$, define

$$f^{-1}\mathcal{F} = \{f^{-1}B : \ B \in \mathcal{F}\}$$

where $f^{-1}B$ is as defined in 2.1. Show that $f^{-1}\mathcal{F}$ is a σ-algebra on E. It is the smallest σ-algebra on E such that f is measurable relative to it and $\mathcal{F}$. It is called the σ-algebra generated by f. If $(E, \mathcal{E})$ is a measurable space, then f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ if and only if $f^{-1}\mathcal{F} \subset \mathcal{E}$; this is another way of stating the definition of measurability.

2.21 *Product spaces.* Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, $(G, \mathcal{G})$ be measurable spaces. Let $f : E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and let $g : E \mapsto G$ be measurable relative to $\mathcal{E}$ and $\mathcal{G}$. Define $h : E \mapsto F \times G$ by

$$h(x) = (f(x), g(x)), \quad x \in E.$$

Show that h is measurable relative to $\mathcal{E}$ and $\mathcal{F} \otimes \mathcal{G}$.

2.22 *Sections.* Let $f : E \times F \mapsto G$ be measurable relative to $\mathcal{E} \otimes \mathcal{F}$ and $\mathcal{G}$. Show that, for fixed x_0 in E, the mapping $h : y \mapsto f(x_0, y)$ is measurable relative to $\mathcal{F}$ and $\mathcal{G}$. (Hint: Note that $h = f \circ g$ where $g : F \mapsto E \times F$ is defined by $g(y) = (x_0, y)$ and show that g is measurable relative to $\mathcal{F}$ and $\mathcal{E} \otimes \mathcal{F}$.) The mapping h is called the *section* of f at x_0.

2.23 *Proposition 2.9.* Prove.

2.24 *Discrete spaces.* Suppose that E is countable and $\mathcal{E} = 2^E$, the discrete σ-algebra on E. Then, $(E, \mathcal{E})$ is said to be *discrete.* Show that every function on E is $\mathcal{E}$-measurable.

2.25 Suppose that $\mathcal{E}$ is generated by a countable partition of E. Show that, then, a numerical function on E is $\mathcal{E}$-measurable if and only if it is constant over each member of that partition.

2.26 *Elementary functions.* A function f on E is said to be *elementary* if it has the form

$$f = \sum_1^\infty a_i 1_{A_i},$$

where $a_i \in \bar{\mathbb{R}}$ and $A_i \in \mathcal{E}$ for each i, the A_i being disjoint. Show that every such function is $\mathcal{E}$-measurable.

2.27 *Measurable functions.* Show that a positive function f on E is $\mathcal{E}$-measurable if and only if it has the form

$$f = \sum_1^\infty a_n 1_{A_n},$$

for some sequence $(a_n) \subset \bar{\mathbb{R}}_+$ and some sequence $(A_n) \subset \mathcal{E}$, disjointedness not required.

2.28 *Approximation by simple functions.* Show that a numerical function f on E is $\mathcal{E}$-measurable if and only if it is the limit of a sequence (f_n) of simple functions. Hint: For necessity, put $f_n = f_n^+ - f_n^-$, where $f_n^+ = d_n \circ f^+$ and $f_n^- = d_n \circ f^-$ with d_n as in Lemma 2.16.

2.29 *Arithmetic operations.* Let f and g be $\mathcal{E}$-measurable. Show that, then, each one of

$$f + g, \qquad f - g, \qquad f \cdot g, \qquad f/g$$

is $\mathcal{E}$-measurable provided that it be well-defined (the issue arises from the fact that $+\infty - \infty$, $(+\infty)(-\infty)$, $0/0$, ∞/∞ are undefined). Recall, however, that $0 \cdot \infty = 0$ is defined.

2.30 *Continuous functions.* Suppose that E is topological. Show that every continuous function $f : E \mapsto \bar{\mathbb{R}}$ is a Borel function. Hint: If f is continuous, then $f^{-1}B$ is open for every open subset of $\bar{\mathbb{R}}$.

2.31 *Step functions, right-continuous functions.* a) A function $f : \mathbb{R}_+ \mapsto \bar{\mathbb{R}}$ is said to be a right-continuous step function if there is a sequence (t_n) in $\mathbb{R}_+$ with $0 = t_0 < t_1 < \cdots$ and $\lim t_n = +\infty$ such that f is constant over each interval $[t_n, t_{n+1})$. Every such function is elementary and, thus, Borel measurable. b) Let $f : \mathbb{R}_+ \mapsto \bar{\mathbb{R}}$ be right-continuous, that is, $f(r_n) \to f(r)$ whenever (r_n) is a sequence decreasing to r. Show that f is Borel measurable. Hint: Note that $f = \lim f_n$, where $f_n = f \circ \bar{d}_n$ for n in $\mathbb{N}^*$ with

$$\bar{d}_n(r) = \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{[\frac{k-1}{2^n}, \frac{k}{2^n})}(r), \quad r \in \mathbb{R}_+.$$

Extend this to $f : \mathbb{R} \mapsto \bar{\mathbb{R}}$ by symmetry on $\mathbb{R} \setminus \mathbb{R}_+$. Similarly, every left-continuous function is Borel.

2.32 *Increasing functions.* Let $f : \mathbb{R} \mapsto \bar{\mathbb{R}}$ be increasing. Show that f is Borel measurable.

2.33 *Measurability of sets defined by functions.* We introduce the notational principle that $\{f \in B\}$, $\{f > r\}$, $\{f \le g\}$, etc. stand for, respectively,

$$\{x \in E : \ f(x) \in B\}, \quad \{x \in E : \ f(x) > r\}, \quad \{x \in E : \ f(x) \le g(x)\},$$

etc. For instance, $\{f \le g\}$ is the set on which f is dominated by g.

Let f and g be $\mathcal{E}$-measurable functions on E. Show that the following sets are in $\mathcal{E}$:

$$\{f > g\}, \quad \{f < g\}, \quad \{f \ne g\}, \quad \{f = g\}, \quad \{f \ge g\}, \quad \{f \le g\}.$$

Hint: $\{f > g\}$ is the set of all x for which $f(x) > r$ and $g(x) < r$ for some rational number r.

2.34 *Positive monotone classes.* This is a variant of the monotone class theorem 2.19: Let $\mathcal{M}_+$ be a collection of positive functions on E. Suppose that

a) $1 \in \mathcal{M}_+$
b) $f, g \in \mathcal{M}_+$ and $a, b \in \mathbb{R}$ and $af + bg \ge 0 \ \Rightarrow \ af + bg \in \mathcal{M}_+$
c) $(f_n) \subset \mathcal{M}_+, \ f_n \nearrow f \ \Rightarrow \ f \in \mathcal{M}_+$.

Suppose, for some p-system $\mathcal{C}$ generating $\mathcal{E}$ that $1_A \in \mathcal{M}_+$ for each A in $\mathcal{C}$. Then, $\mathcal{M}_+$ includes every positive $\mathcal{E}$-measurable function. Prove.

2.35 *Bounded monotone classes.* This is another variant of the monotone class theorem. Let $\mathcal{M}_b$ be a collection of bounded functions on E. Suppose that

a) $1 \in \mathcal{M}_b$,
b) $f, g \in \mathcal{M}_b$ and $a, b \in \mathbb{R} \Rightarrow af + bg \in \mathcal{M}_b$,
c) $(f_n) \subset \mathcal{M}_b, f_n \ge 0, f_n \nearrow f,$ and f is bounded $\Rightarrow f \in \mathcal{M}_b$.

Suppose, for some p-system $\mathcal{C}$ generating $\mathcal{E}$ that $1_A \in \mathcal{M}_b$ for each A in $\mathcal{C}$. Then, $\mathcal{M}_b$ includes every bounded $\mathcal{E}$-measurable function. Prove.

3 Measures

Let $(E, \mathcal{E})$ be a measurable space, that is, E is a set and $\mathcal{E}$ is a σ-algebra on E. A *measure* on $(E, \mathcal{E})$ is a mapping $\mu : \mathcal{E} \mapsto \bar{\mathbb{R}}_+$ such that

3.1 a) $\mu(\emptyset) = 0$,
b) $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ for every disjointed sequence (A_n) in $\mathcal{E}$.

The latter condition is called *countable additivity*. Note that $\mu(A)$ is always positive and can be $+\infty$; the number $\mu(A)$ is called the measure of A; we also write μA for it.

A *measure space* is a triplet $(E, \mathcal{E}, \mu)$, where $(E, \mathcal{E})$ is a measurable space and μ is a measure on it.

Examples

3.2 *Dirac measures.* Let $(E, \mathcal{E})$ be a measurable space, and let x be a fixed point of E. For each A in $\mathcal{E}$, put

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Then, δ_x is a measure on $(E, \mathcal{E})$. It is called the *Dirac measure* sitting at x.

3.3 *Counting measures.* Let $(E, \mathcal{E})$ be a measurable space. Let D be a fixed subset of E. For each A in $\mathcal{E}$, let $\nu(A)$ be the number of points in $A \cap D$. Then, ν is a measure on $(E, \mathcal{E})$. Such ν are called *counting measures.* Often, the set D is taken to be countable, in which case

$$\nu(A) = \sum_{x \in D} \delta_x(A), \quad A \in \mathcal{E}.$$

3.4 *Discrete measures.* Let $(E, \mathcal{E})$ be a measurable space. Let D be a countable subset of E. For each x in D, let $m(x)$ be a positive number. Define

$$\mu(A) = \sum_{x \in D} m(x)\, \delta_x(A), \quad A \in \mathcal{E}.$$

Then, μ is a measure on $(E, \mathcal{E})$. Such measures are said to be *discrete.* We may think of $m(x)$ as the mass attached to the point x, and then $\mu(A)$ is the mass on the set A. In particular, if $(E, \mathcal{E})$ is a discrete measurable space, then every measure μ on it has this form.

3.5 *Lebesgue measures.* A measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is called the Lebesgue measure on $\mathbb{R}$ if $\mu(A)$ is the length of A for every interval A. As with most measures, it is impossible to display $\mu(A)$ for every Borel set A, but one can do integration with it, which is the main thing measures are for. Similarly, the Lebesgue measure on $\mathbb{R}^2$ is the "area" measure, on $\mathbb{R}^3$ the "volume", etc. We shall write *Leb* for them. Also note the harmless vice of saying, for example, Lebesgue measure on $\mathbb{R}^2$ to mean Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

Some properties

3.6 Proposition. *Let μ be a measure on a measurable space $(E, \mathcal{E})$. Then, the following hold for all measurable sets A, B, and $A_1, A_2, \ldots$:*

Finite additivity: $A \cap B = \emptyset \;\Rightarrow\; \mu(A \cup B) = \mu(A) + \mu(B)$.

Monotonicity: $A \subset B \;\Rightarrow\; \mu(A) \le \mu(B)$.

Sequential continuity: $A_n \nearrow A \;\Rightarrow\; \mu(A_n) \nearrow \mu(A)$.

Boole's inequality: $\mu(\bigcup_n A_n) \le \sum_n \mu(A_n)$.

Proof. Finite additivity is a particular instance of countable additivity of μ: take $A_1 = A$, $A_2 = B$, $A_3 = A_4 = \ldots = \emptyset$ in 3.1b. Monotonicity follows from finite additivity and the positivity of μ: for $A \subset B$, we can write B as the union of disjoint sets A and $B \setminus A$, and hence

$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A),$$

since $\mu(B \setminus A) \ge 0$. Sequential continuity follows from countable additivity: Suppose that $A_n \nearrow A$. Then, $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2, \ldots$ are disjoint, their union is A, and the union of the first n is A_n. Thus, the sequence of numbers $\mu(A_n)$ increases and

$$\lim \mu(A_n) = \lim \mu(\cup_1^n B_i) = \lim \sum_1^n \mu(B_i) = \sum_1^\infty \mu(B_i) = \mu(A).$$

Finally, to show Boole's inequality, we start by observing that

$$\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \;\le\; \mu(A) + \mu(B)$$

for arbitrary A and B in $\mathcal{E}$. This extends to finite unions by induction:

$$\mu(\cup_1^n A_i) \;\le\; \sum_1^n \mu(A_i).$$

Taking limits on both sides completes the proof since the left side has limit $\mu(\cup_1^\infty A_i)$ by sequential continuity. □

Arithmetic of measures

Let $(E, \mathcal{E})$ be a measurable space. If μ is a measure on it and $c > 0$ is a constant, then $c\mu$ is again a measure on it. If μ and ν are measures on it, then so is $\mu + \nu$. If $\mu_1, \mu_2, \ldots$ are measures, then so is $\sum_n \mu_n$; this can be checked using the elementary fact that, if the numbers a_{mn} are positive,

$$\sum_m \sum_n a_{mn} = \sum_n \sum_m a_{mn}.$$

Finite, σ-finite, Σ-finite measures

Let μ be a measure on a measurable space $(E, \mathcal{E})$. It is said to be *finite* if $\mu(E) < \infty$; then $\mu(A) < \infty$ for all A in $\mathcal{E}$ by the monotonicity of μ. It is called a *probability measure* if $\mu(E) = 1$. It is said to be σ*-finite* if there exists a measurable partition (E_n) of E such that $\mu(E_n) < \infty$ for each n. Finally, it is said to be Σ-finite if there exists a sequence of finite measures μ_n such that $\mu = \sum_n \mu_n$. Every finite measure is obviously σ-finite. Every σ-finite measure is Σ-finite; see Exercise 3.13 for this point and for examples.

Specification of measures

Given a measure on $(E, \mathcal{E})$, its values over a p-system generating $\mathcal{E}$ determine its values over all of $\mathcal{E}$, generally. The following is the precise statement for finite measures. Its version for σ-finite measures is given in Exercise 3.18.

3.7 Proposition. *Let $(E, \mathcal{E})$ be a measurable space. Let μ and ν be measures on it with $\mu(E) = \nu(E) < \infty$. If μ and ν agree on a p-system generating $\mathcal{E}$, then μ and ν are identical.*

Proof. Let $\mathcal{C}$ be a p-system generating $\mathcal{E}$. Suppose that $\mu(A) = \nu(A)$ for every A in $\mathcal{C}$, and $\mu(E) = \nu(E) < \infty$. We need to show that, then, $\mu(A) = \nu(A)$ for every A in $\mathcal{E}$, or equivalently, that

$$\mathcal{D} = \{A \in \mathcal{E} : \ \mu(A) = \nu(A)\}$$

contains $\mathcal{E}$. Since $\mathcal{D} \supset \mathcal{C}$ by assumption, it is enough to show that $\mathcal{D}$ is a d-system, for, then, the monotone class theorem 1.8 yields the desired conclusion that $\mathcal{D} \supset \mathcal{E}$. So, we check the conditions for $\mathcal{D}$ to be a d-system. First, $E \in \mathcal{D}$ by the assumption that $\mu(E) = \nu(E)$. If $A, B \in \mathcal{D}$, and $A \supset B$, then $A \setminus B \in \mathcal{D}$, because

$$\mu(A \setminus B) = \mu(A) - \mu(B) = \nu(A) - \nu(B) = \nu(A \setminus B),$$

where we used the finiteness of μ to solve $\mu(A) = \mu(B) + \mu(A \setminus B)$ for $\mu(A \setminus B)$ and similarly for $\nu(A \setminus B)$. Finally, suppose that $(A_n) \subset \mathcal{D}$ and $A_n \nearrow A$; then, $\mu(A_n) = \nu(A_n)$ for every n, the left side increases to $\mu(A)$ by the sequential continuity of μ, and the right side to $\nu(A)$ by the same for ν; hence, $\mu(A) = \nu(A)$ and $A \in \mathcal{D}$. □

3.8 Corollary. *Let μ and ν be probability measures on $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$. Then, $\mu = \nu$ if and only if $\mu[-\infty, r] = \nu[-\infty, r]$ for every r in $\mathbb{R}$.*

Proof is immediate from the preceding proposition: $\mu(\bar{\mathbb{R}}) = \nu(\bar{\mathbb{R}}) = 1$ since μ and ν are probability measures, and the intervals $[-\infty, r]$ with r in $\mathbb{R}$ form a p-system generating the Borel σ-algebra on $\bar{\mathbb{R}}$.

Atoms, purely atomic measures, diffuse measures

Let $(E, \mathcal{E})$ be a measurable space. Suppose that the singleton $\{x\}$ belongs to $\mathcal{E}$ for every x in E; this is true for all standard measurable spaces. Let μ be a measure on $(E, \mathcal{E})$. A point x is said to be an *atom* of μ if $\mu\{x\} > 0$. The measure μ is said to be *diffuse* if it has no atoms. It is said to be *purely atomic* if the set D of its atoms is countable and $\mu(E \setminus D) = 0$. For example, Lebesgue measures are diffuse, a Dirac measure is purely atomic with one atom, discrete measures are purely atomic.

The following proposition applies to Σ-finite (and therefore, to finite and σ-finite) measures. We leave the proof as an exercise; see 3.15.

3.9 PROPOSITION. *Let μ be a Σ-finite measure on $(E, \mathcal{E})$. Then,*

$$\mu = \lambda + \nu,$$

where λ is a diffuse measure and ν is purely atomic.

Completeness, negligible sets

Let $(E, \mathcal{E}, \mu)$ be a measure space. A measurable set B is said to be *negligible* if $\mu(B) = 0$. An arbitrary subset of E is said to be *negligible* if it is contained in a measurable negligible set. The measure space is said to be *complete* if every negligible set is measurable. If it is not complete, the following shows how to enlarge $\mathcal{E}$ to include all negligible sets and to extend μ onto the enlarged $\mathcal{E}$. We leave the proof to Exercise 3.16. The measure space $(E, \bar{\mathcal{E}}, \bar{\mu})$ described is called the *completion* of $(E, \mathcal{E}, \mu)$. When $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$ and $\mu = Leb$, the elements of $\bar{\mathcal{E}}$ are called the *Lebesgue measurable* sets.

3.10 PROPOSITION. *Let $\mathcal{N}$ be the collection of all negligible subsets of E. Let $\bar{\mathcal{E}}$ be the σ-algebra generated by $\mathcal{E} \cup \mathcal{N}$. Then,*

a) every B in $\bar{\mathcal{E}}$ has the form $B = A \cup N$, where $A \in \mathcal{E}$ and $N \in \mathcal{N}$,

b) the formula $\bar{\mu}(A \cup N) = \mu(A)$ defines a unique measure $\bar{\mu}$ on $\bar{\mathcal{E}}$, we have $\bar{\mu}(A) = \mu(A)$ for $A \in \mathcal{E}$, and the measure space $(E, \bar{\mathcal{E}}, \bar{\mu})$ is complete.

Almost everywhere

If a proposition holds for all but a negligible set of x in E, then we say that it holds for almost every x, or almost everywhere. If the measure μ used to define negligibility needs to be indicated, we say μ-almost every x or μ-almost everywhere. If E is replaced by a measurable set A, we say almost everywhere on A. For example, given numerical functions f and g on E, and a measurable set A, saying that $f = g$ almost everywhere on A is equivalent to saying that $\{x \in A : f(x) \neq g(x)\}$ is negligible, which is then equivalent to saying that there exists a measurable set M with $\mu(M) = 0$ such that $f(x) = g(x)$ for every x in $A \setminus M$.

Exercises and complements

3.11 *Restrictions and traces.* Let $(E, \mathcal{E})$ be a measurable space, and μ a measure on it. Let $D \in \mathcal{E}$.

a) Define $\nu(A) = \mu(A \cap D)$, $A \in \mathcal{E}$. Show that ν is a measure on $(E, \mathcal{E})$; it is called the trace of μ on D.

b) Let $\mathcal{D}$ be the trace of $\mathcal{E}$ on D (see 1.15). Define $\nu(A) = \mu(A)$ for A in $\mathcal{D}$. Show that ν is a measure on $(D, \mathcal{D})$; it is called the restriction of μ to D.

3.12 *Extensions.* Let $(E, \mathcal{E})$ be a measurable space, let $D \in \mathcal{E}$, and let $(D, \mathcal{D})$ be the trace of $(E, \mathcal{E})$ on D. Let μ be a measure on $(D, \mathcal{D})$ and define ν by

$$\nu(A) = \mu(A \cap D), \quad A \in \mathcal{E}.$$

Show that ν is a measure on $(E, \mathcal{E})$. This device allows us to regard a "measure on D" as a "measure on E".

3.13 *σ-and Σ-finiteness*

a) Let $(E, \mathcal{E})$ be a measurable space. Let μ be a σ-finite measure on it. Then, μ is Σ-finite. Show. Hint: Let (E_n) be a measurable partition of E such that $\mu(E_n) < \infty$ for each n; define μ_n to be the trace of μ on E_n as in Exercise 3.11a; show that $\mu = \sum_n \mu_n$.

b) Show that the Lebesgue measure on $\mathbb{R}$ is σ-finite.

c) Let μ be the discrete measure of Example 3.4 with $(E, \mathcal{E})$ discrete. Show that it is σ-finite if and only if $m(x) < \infty$ for every x in D. Show that it is always Σ-finite.

d) Let $E = [0, 1]$ and $\mathcal{E} = \mathcal{B}(E)$. For A in $\mathcal{E}$, define $\mu(A)$ to be 0 if $\text{Leb}\, A = 0$ and $+\infty$ if $\text{Leb}\, A > 0$. Show that μ is not σ-finite but is Σ-finite.

e) Let $(E, \mathcal{E})$ be as in (d) here. Define $\mu(A)$ to be the counting measure on it (see Example 3.3 and take $D = E$). Show that μ is neither σ-finite nor Σ-finite.

3.14 *Atoms.* Show that a finite measure has at most countably many atoms. Show that the same is true for Σ-finite measures. Hint: If $\mu(E) < \infty$ then the number of atoms with $\mu\{x\} > \frac{1}{n}$ is at most $n\mu(E)$.

3.15 *Proof of Proposition 3.9.* Let D be the set of all atoms of the given Σ-finite measure μ. Then, D is countable by the preceding exercise and, thus, measurable by the measurability of singletons. Define

$$\lambda(A) = \mu(A \setminus D), \qquad \nu(A) = \mu(A \cap D), \qquad A \in \mathcal{E}.$$

Show that λ is a diffuse measure, ν purely atomic, and $\mu = \lambda + \nu$. Note that ν has the form in Example 3.4 with $m(x) = \mu\{x\}$ for each atom x.

3.16 *Proof of Proposition 3.10.* Let $\mathcal{F}$ be the collection of all sets having the form $A \cup N$ with A in $\mathcal{E}$ and N in $\mathcal{N}$. Show that $\mathcal{F}$ is a σ-algebra on E. Argue that $\mathcal{F} = \bar{\mathcal{E}}$, thus proving part (a). To show (b), we need to show

that, if $A \cup N = A' \cup N'$ with A and A' in $\mathcal{E}$ and N and N' in $\mathcal{N}$, then $\mu(A) = \mu(A')$. To this end pick M in $\mathcal{E}$ such that $\mu(M) = 0$ and $M \supset N$, and pick M' similarly for N'. Show that $A \subset A' \cup M'$ and $A' \subset A \cup M$. Use this, monotonicity of μ, Boole's inequality, etc. several times to show that $\mu(A) = \mu(A')$.

3.17 *Measurability on completions.* Let $(E, \mathcal{E}, \mu)$ be a measure space, and $(E, \bar{\mathcal{E}}, \bar{\mu})$ its completion. Let f be a numerical function on E. Show that f is $\bar{\mathcal{E}}$-measurable if and only if there exists an $\mathcal{E}$-measurable function g such that $f = g$ almost everywhere. Hint: For sufficiency, choose M in $\mathcal{E}$ such that $\mu(M) = 0$ and $f = g$ outside M, and note that $\{f \le r\} = A \cup N$ where $A = \{g \le r\} \setminus M$ and $N \subset M$. For necessity, assuming f is positive $\bar{\mathcal{E}}$-measurable, write $f = \sum_1^\infty a_n 1_{A_n}$ with $A_n \in \bar{\mathcal{E}}$ for each n (see Exercise 2.27) and choosing B_n in $\mathcal{E}$ such that $A_n = B_n \cup N_n$ for some negligible N_n, define $g = \sum_1^\infty a_n 1_{B_n}$, and show that $\{f \ne g\} \subset \bigcup_n N_n = N$, which is negligible.

3.18 *Equality of measures.* This is to extend Proposition 3.7 to σ-finite measures. Let μ and ν be such measures on $(E, \mathcal{E})$. Suppose that they agree on a p-system $\mathcal{C}$ that generates $\mathcal{E}$. Suppose further that $\mathcal{C}$ contains a partition (E_n) of E such that $\mu(E_n) = \nu(E_n) < \infty$ for each n. Then, $\mu = \nu$. Prove this.

3.19 *Existence of probability measures.* Let E be a set, $\mathcal{D}$ an algebra on it, and put $\mathcal{E} = \sigma\mathcal{D}$. Suppose that $\lambda : \mathcal{D} \mapsto [0, 1]$ is such that $\lambda(E) = 1$ and $\lambda(A \cup B) = \lambda(A) + \lambda(B)$ whenever A and B are disjoint sets in $\mathcal{D}$. Is it possible to extend λ to a probability measure on $\mathcal{E}$? In other words, does there exist a measure μ on $(E, \mathcal{E})$ such that $\mu(A) = \lambda(A)$ for every A in $\mathcal{D}$? If such a measure exists, then it is unique by Proposition 3.7, since $\mathcal{D}$ is a p-system that generates $\mathcal{E}$.
The answer is provided by *Caratheodory's extension theorem*, a classical result. Such a probability measure μ exists provided that λ be countably additive on $\mathcal{D}$, that is, if (A_n) is a disjointed sequence in $\mathcal{D}$ with $A = \bigcup_n A_n \in \mathcal{D}$, then we must have $\lambda(A) = \sum_n \lambda(A_n)$, or equivalently, if $(A_n) \subset \mathcal{D}$ and $A_n \searrow \emptyset$ then we must have $\lambda(A_n) \searrow 0$.

4 Integration

Let $(E, \mathcal{E}, \mu)$ be a measure space. Recall that $\mathcal{E}$ stands also for the collection of all $\mathcal{E}$-measurable functions on E and that $\mathcal{E}_+$ is the sub-collection consisting of positive $\mathcal{E}$-measurable functions. Our aim is to define the "integral of f with respect to μ" for all reasonable functions f in $\mathcal{E}$. We shall denote it by any of the following:

4.1
$$\mu f = \mu(f) = \int_E \mu(dx) f(x) = \int_E f \, d\mu.$$

As the notation μf suggests, integration is a kind of multiplication; this will become clear when we show that the following hold for all a, b in $\mathbb{R}_+$ and f, g, f_n in $\mathcal{E}_+$:

4.2 a) *Positivity*: $\mu f \geq 0$; $\mu f = 0$ if $f = 0$.
b) *Linearity*: $\mu(af + bg) = a\,\mu f + b\,\mu g$.
c) *Monotone convergence theorem*: If $f_n \nearrow f$, then $\mu f_n \nearrow \mu f$.

We start with the definition of the integral and proceed to proving the properties 4.2 and their extensions. At the end, we shall also show that 4.2 characterizes integration.

4.3 Definition. *a) Let f be simple and positive. If its canonical form is $f = \sum_1^n a_i 1_{A_i}$, then we define*

$$\mu f = \sum_1^n a_i\, \mu(A_i).$$

b) Let $f \in \mathcal{E}_+$. Put $f_n = d_n \circ f$, where the d_n are as in Lemma 2.16. Then each f_n is simple and positive, and the sequence (f_n) increases to f as shown in the proof of 2.17. The integral μf_n is defined for each n by the preceding step, and the sequence of numbers μf_n is increasing (see Remark 4.4d below). We define

$$\mu f = \lim \mu f_n.$$

c) Let $f \in \mathcal{E}$. Then, $f^+ = f \vee 0$ and $f^- = -(f \wedge 0)$ belong to $\mathcal{E}_+$, and their integrals $\mu(f^+)$ and $\mu(f^-)$ are defined by the preceding step. Noting that $f = f^+ - f^-$, we define

$$\mu f = \mu(f^+) - \mu(f^-)$$

provided that at least one term on the right side be finite. Otherwise, if $\mu(f^+) = \mu(f^-) = +\infty$, then μf is undefined.

4.4 Remarks. Let f, g, etc. be simple and positive.

a) The formula for μf remains the same even when $f = \sum a_i 1_{A_i}$ is not the canonical representation of f. This is easy to check using the finite additivity of μ.

b) If a and b are in $\mathbb{R}_+$, then $af + bg$ is simple and positive, and the linearity property holds:

$$\mu(af + bg) = a\,\mu f + b\,\mu g.$$

This can be checked using the preceding remark.

c) If $f \leq g$ then $\mu f \leq \mu g$. This follows from the linearity property above applied to the simple positive functions f and $g - f$:

$$\mu f \leq \mu f + \mu(g - f) = \mu(f + g - f) = \mu g.$$

d) In step (b) of the definition, we have $f_1 \leq f_2 \leq \ldots$. The preceding remark on monotonicity shows that $\mu f_1 \leq \mu f_2 \leq \ldots$. Thus, $\lim \mu f_n$ exists as claimed (it can be $+\infty$).

Examples

a) *Discrete measures.* Fix x_0 in E and consider the Dirac measure δ_{x_0} sitting at x_0. Going through the steps of the definition of the integral, we see that $\delta_{x_0} f = f(x_0)$ for every f in $\mathcal{E}$. This extends to discrete measures: if $\mu = \sum_{x \in D} m(x)\delta_x$ for some countable set D and positive masses $m(x)$, then

$$\mu f = \sum_{x \in D} m(x)\ f(x)$$

for every f in $\mathcal{E}_+$. A similar result holds for purely atomic measures as well.

b) *Discrete spaces.* Suppose that $(E, \mathcal{E})$ is discrete, that is, E is countable and $\mathcal{E} = 2^E$. Then, every numerical function on E is $\mathcal{E}$-measurable, and every measure μ has the form in the preceding example with $D = E$ and $m(x) = \mu\{x\}$. Thus, for every positive function f on E,

$$\mu f = \sum_{x \in E} \mu\{x\} f(x).$$

In this case, and especially when E is finite, every function can be thought as a vector, and similarly for every measure. Further, we think of functions as column vectors and of measures as row vectors. Then, the integral μf is seen to be the product of the row vector μ and the column vector f. So, the notation is well-chosen in this case and extends to arbitrary spaces in a most suggestive manner.

c) *Lebesgue integrals.* Suppose that E is a Borel subset of $\mathbb{R}^d$ for some $d \geq 1$ and suppose that $\mathcal{E} = \mathcal{B}(E)$, the Borel subsets of E. Suppose that μ is the restriction of the Lebesgue measure on $\mathbb{R}^d$ to $(E, \mathcal{E})$. For f in $\mathcal{E}$, we employ the following notations for the integral μf:

$$\mu f = \text{Leb}_E\ f = \int_E \text{Leb}(dx)\ f(x) = \int_E dx\ f(x),$$

the last using dx for $\text{Leb}(dx)$ in keeping with tradition. This integral is called the Lebesgue integral of f on E.

If the Riemann integral of f exists, then so does the Lebesgue integral, and the two integrals are equal. The converse is false; the Lebesgue integral exists for a larger class of functions than does the Riemann integral. For example, if $E = [0, 1]$, and f is the indicator of the set of all rational numbers in E, then the Lebesgue integral of f is well-defined by 4.3a to be zero, but the Riemann integral does not exist because the discontinuity set of f in E is E itself and $\text{Leb}\, E = 1 \neq 0$ (recall that a Borel function is Riemann integrable over an interval $[a, b]$ if and only if its points of discontinuity in $[a, b]$ form a set of Lebesgue measure 0).

Integrability

A function f in $\mathcal{E}$ is said to be *integrable* if μf exists and is a real number. Thus, f in $\mathcal{E}$ is integrable if and only if $\mu f^+ < \infty$ and $\mu f^- < \infty$, or

equivalently, if and only if the integral of $|f| = f^+ + f^-$ is a finite number. We leave it as an exercise to show that every integrable function is real-valued almost everywhere.

Integral over a set

Let $f \in \mathcal{E}$ and let A be a measurable set. Then, $f1_A \in \mathcal{E}$, and the *integral of f over A* is defined to be the integral of $f1_A$. The following notations are used for it:

4.5 $$\mu(f1_A) = \int_A \mu(dx)f(x) = \int_A f\,d\mu.$$

The following shows that, for each f in $\mathcal{E}_+$, the set function $A \mapsto \mu(f1_A)$ is finitely additive. This property extends to countable additivity as a corollary to the monotone convergence theorem 4.8 below.

4.6 LEMMA. *Let $f \in \mathcal{E}_+$. Let A and B be disjoint sets in $\mathcal{E}$ with union C. Then*

$$\mu(f1_A) + \mu(f1_B) = \mu(f1_C).$$

Proof. If f is simple, this is immediate from the linearity property of Remark 4.4b. For arbitrary f in $\mathcal{E}_+$, putting $f_n = d_n \circ f$ as in Definition 4.3b, we get

$$\mu(f_n1_A) + \mu(f_n1_B) = \mu(f_n1_C)$$

since the f_n are simple. Observing that $f_n1_D = d_n \circ (f1_D)$ for $D = A, B, C$ and taking limits as $n \to \infty$ we get the desired result through Definition 4.3b. □

Positivity and monotonicity

4.7 PROPOSITION. *If $f \in \mathcal{E}_+$, then $\mu f \geq 0$. If f and g are in $\mathcal{E}_+$ and $f \leq g$, then $\mu f \leq \mu g$.*

Proof. Positivity of μf for f positive is immediate from Definition 4.3. To show monotonicity, let $f_n = d_n \circ f$ and $g_n = d_n \circ g$ as in step 4.3b. Since each d_n is an increasing function (see Lemma 2.16), $f \leq g$ implies that $f_n \leq g_n$ for each n which in turn implies that $\mu f_n \leq \mu g_n$ for each n by Remark 4.4c. Letting $n \to \infty$, we see from Definition 4.3b that $\mu f \leq \mu g$. □

Monotone Convergence Theorem

This is the main theorem of integration. It is the key tool for interchanging the order of taking limits and integrals. It states that the mapping $f \mapsto \mu f$ from $\mathcal{E}_+$ into $\bar{\mathbb{R}}_+$ is continuous under increasing limits. As such, it is an extension of the sequential continuity of measures.

4.8 THEOREM. *Let (f_n) be an increasing sequence in $\mathcal{E}_+$. Then,*

$$\mu(\lim f_n) = \lim \mu f_n.$$

Proof. Let $f = \lim f_n$; it is well-defined since (f_n) is increasing. Clearly, $f \in \mathcal{E}_+$, and μf is well-defined. Since (f_n) is increasing, the integrals μf_n form an increasing sequence of numbers by the monotonicity property shown by Proposition 4.7. Hence, $\lim \mu f_n$ exists. We want to show that the limit is μf. Since $f \geq f_n$ for each n, we have $\mu f \geq \mu f_n$ by the monotonicity property. It follows that $\mu f \geq \lim \mu f_n$. The following steps show that the reverse inequality holds as well.

a) Fix b in $\mathbb{R}_+$ and B in $\mathcal{E}$. Suppose that $f(x) > b$ for every x in the set B. Since the sets $\{f_n > b\}$ are increasing to $\{f > b\}$, the sets $B_n = B \cap \{f_n > b\}$ are increasing to B, and

4.9 $$\lim \mu(B_n) = \mu(B)$$

by the sequential continuity of μ. On the other hand,

$$f_n 1_B \geq f_n 1_{B_n} \geq b 1_{B_n},$$

which yields via monotonicity that

$$\mu(f_n 1_B) \geq \mu(b 1_{B_n}) = b\mu(B_n).$$

Taking note of 4.9 we conclude that

4.10 $$\lim \mu(f_n 1_B) \geq b\mu(B).$$

This remains true if $f(x) \geq b$ for all x in B: If $b = 0$ then this is trivially true. If $b > 0$ then choose a sequence (b_m) strictly increasing to b; then, 4.10 holds with b replaced by b_m; and letting $m \to \infty$ we obtain 4.10 again.

b) Let g be a positive simple function such that $f \geq g$. If $g = \sum_1^m b_i 1_{B_i}$ is its canonical representation, then $f(x) \geq b_i$ for every x in B_i, and 4.10 yields

$$\lim_n \mu(f_n 1_{B_i}) \geq b_i \mu(B_i), \qquad i = 1, \ldots, m.$$

Hence, by the finite additivity of $A \mapsto \mu(f_n 1_A)$ shown in Lemma 4.6,

4.11 $$\lim_n \mu f_n = \lim_n \sum_{i=1}^m \mu(f_n 1_{B_i}) = \sum_{i=1}^m \lim_n \mu(f_n 1_{B_i}) \geq \sum_{i=1}^m b_i \mu(B_i) = \mu g.$$

c) Recall that $\mu f = \lim \mu(d_k \circ f)$ by Definition 4.3b. For each k, the function $d_k \circ f$ is simple and $f \geq d_k \circ f$. Hence, taking $g = d_k \circ f$ in 4.11, we have

$$\lim_n \mu f_n \geq \mu(d_k \circ f)$$

for all k. Letting $k \to \infty$ we obtain the desired inequality that $\lim \mu f_n \geq \mu f$. □

Linearity of integration

4.12 Proposition. *For f and g in $\mathcal{E}_+$ and a and b in $\mathbb{R}_+$,*

$$\mu(af + bg) = a\,\mu f + b\,\mu g.$$

The same is true for integrable f and g in $\mathcal{E}$ and arbitrary a and b in $\mathbb{R}$.

Proof. Suppose that f, g, a, b are all positive. If f and g are simple, the linearity can be checked directly as remarked in 4.4b. If not, choose (f_n) and (g_n) to be sequences of simple positive functions increasing to f and g respectively. Then,

$$\mu(af_n + bg_n) = a\,\mu f_n + b\,\mu g_n,$$

and the monotone convergence theorem applied to both sides completes the proof. The remaining statements follow from Definition 4.3c and the linearity for positive functions after putting $f = f^+ - f^-$ and $g = g^+ - g^-$. □

Insensitivity of the integral

We show next that the integral of a function remains unchanged if the values of the function are changed over a negligible set.

4.13 Proposition. *If A in $\mathcal{E}$ is negligible, then $\mu(f1_A) = 0$ for every f in $\mathcal{E}$. If f and g are in $\mathcal{E}_+$ and $f = g$ almost everywhere, then $\mu f = \mu g$. If $f \in \mathcal{E}_+$ and $\mu f = 0$, then $f = 0$ almost everywhere.*

Proof. a) Let A be measurable and negligible. If $f \in \mathcal{E}_+$ and simple, then $\mu(f1_A) = 0$ by Definition 4.3a. This extends to the non-simple case by the monotone convergence theorem using a sequence of simple f_n increasing to f: then $\mu(f_n 1_A) = 0$ for all n and $\mu(f1_A)$ is the limit of the left side. For f in $\mathcal{E}$ arbitrary, we have $\mu(f^+1_A) = 0$ and $\mu(f^-1_A) = 0$ and hence $\mu(f1_A) = 0$ since $(f1_A)^+ = f^+1_A$ and $(f1_A)^- = f^-1_A$.

b) If f and g are in $\mathcal{E}_+$ and $f = g$ almost everywhere, then $A = \{f \neq g\}$ is measurable and negligible, and the integrals of f and g on A both vanish. Thus, with $B = A^c$, we have $\mu f = \mu(f1_B)$ and $\mu g = \mu(g1_B)$, which imply $\mu f = \mu g$ since $f(x) = g(x)$ for all x in B.

c) Let $f \in \mathcal{E}_+$ and $\mu f = 0$. We need to show that the set $N = \{f > 0\}$ has measure 0. Take a sequence of numbers $\varepsilon_k > 0$ decreasing to 0, let $N_k = \{f > \varepsilon_k\}$, and observe that $N_k \nearrow N$, which implies that $\mu(N_k) \nearrow \mu(N)$ by the sequential continuity of μ. Thus, it is enough to show that $\mu(N_k) = 0$ for every k. This is easy to show: $f \geq \varepsilon_k 1_{N_k}$ implies that $\mu f \geq \varepsilon_k \mu(N_k)$, and since $\mu f = 0$ and $\varepsilon_k > 0$, we must have $\mu(N_k) = 0$. □

Fatou's lemma

We return to the properties of the integral under limits. Next is a useful consequence of the monotone convergence theorem.

4.14 LEMMA. *Let* $(f_n) \subset \mathcal{E}_+$. *Then* $\mu(\liminf f_n) \le \liminf \mu f_n$.

Proof. Define $g_m = \inf_{n \ge m} f_n$ and recall that $\liminf f_n$ is the limit of the increasing sequence (g_m) in $\mathcal{E}_+$. Hence, by the monotone convergence theorem,

$$\mu(\liminf f_n) = \lim \mu g_m.$$

On the other hand, $g_m \le f_n$ for all $n \ge m$, which implies that $\mu g_m \le \mu f_n$ for all $n \ge m$ by the monotonicity of integration, which in turn means that $\mu g_m \le \inf_{n \ge m} \mu f_n$. Hence, as desired,

$$\lim \mu g_m \le \liminf \mu f_n. \qquad \square$$

4.15 COROLLARY. *Let* $(f_n) \subset \mathcal{E}$. *If there is an integrable function* g *such that* $f_n \ge g$ *for every* n, *then*

$$\mu(\liminf f_n) \le \liminf \mu f_n.$$

If there is an integrable function g *such that* $f_n \le g$ *for every* n, *then*

$$\mu(\limsup f_n) \ge \limsup \mu f_n.$$

Proof. Let g be integrable. Then, the complement of the measurable set $A = \{g \in \mathbb{R}\}$ is negligible (see Exercise 4.24 for this). Hence, $f_n 1_A = f_n$ almost everywhere, $g 1_A = g$ almost everywhere, and $g 1_A$ is real-valued. The first statement follows from Fatou's Lemma applied to the well-defined sequence $(f_n 1_A - g 1_A)$ in $\mathcal{E}_+$ together with the linearity and insensitivity of integration. The second statement follows again from Fatou's lemma, now applied to the well-defined sequence $(g 1_A - f_n 1_A)$ in $\mathcal{E}_+$ together with the linearity and insensitivity, and the observation that $\limsup r_n = -\liminf(-r_n)$ for every sequence (r_n) in $\bar{\mathbb{R}}$. $\square$

Dominated convergence theorem

This is the second important tool for interchanging the order of taking limits and integrals. A function f is said to be *dominated* by the function g if $|f| \le g$; note that $g \ge 0$ necessarily. A sequence (f_n) is said to be dominated by g if $|f_n| \le g$ for every n. If so, and if g can be taken to be a finite constant, then (f_n) is said to be *bounded.*

4.16 THEOREM. *Let* $(f_n) \subset \mathcal{E}$. *Suppose that* (f_n) *is dominated by some integrable function* g. *If* $\lim f_n$ *exists, then it is integrable and*

$$\mu(\lim f_n) = \lim \mu f_n.$$

Proof. By assumption, $-g \le f_n \le g$ for every n, and both g and $-g$ are integrable. Thus, both statements of the last corollary apply:

4.17 $$\mu(\liminf f_n) \le \liminf \mu f_n \le \limsup \mu f_n \le \mu(\limsup f_n).$$

If $\lim f_n$ exists, then $\liminf f_n = \limsup f_n = \lim f_n$, and $\lim f_n$ is integrable since it is dominated by g. Hence, the extreme members of 4.17 are finite and equal, and all inequality signs are in fact equalities. □

If (f_n) is bounded, say by the constant b, and if the measure μ is finite, then we can take $g = b$ in the preceding theorem. The resulting corollary is called the *bounded convergence theorem*:

4.18 THEOREM. *Let $(f_n) \subset \mathcal{E}$. Suppose that (f_n) is bounded and μ is finite. If $\lim f_n$ exists, then it is a bounded integrable function and*

$$\mu(\lim f_n) = \lim \mu f_n.$$

Almost everywhere versions

The insensitivity of integration to changes over negligible sets enables us to re-state all the results above by allowing the conditions to fail over negligible sets. We start by extending the definition of integration somewhat.

4.19 CONVENTION. *Let f be a numerical function on E. Suppose that there exists an $\mathcal{E}$-measurable function g such that $f(x) = g(x)$ for almost every x in E. Then, we define the integral μf of f to be the number μg provided that μg is defined. Otherwise, if μg does not exist, μf does not exist either.*

The definition here is without ambiguities: if h is another measurable function such that $f = h$ almost everywhere, then $g = h$ almost everywhere; if μg exists, then so does μh and $\mu g = \mu h$ by the insensitivity property; if μg does not exist, then neither does μh.

In fact, the convention here is one of notation making, almost. Let $g \in \mathcal{E}$ and $f = g$ almost everywhere. Let $(E, \bar{\mathcal{E}}, \bar{\mu})$ be the completion of $(E, \mathcal{E}, \mu)$. Then, $f \in \bar{\mathcal{E}}$ (see Exercise 3.17 for this), and the integral $\bar{\mu} f$ makes sense by Definition 4.3 applied on the measurable space $(E, \bar{\mathcal{E}}, \bar{\mu})$. Since $\mathcal{E} \subset \bar{\mathcal{E}}$, the function g is $\bar{\mathcal{E}}$-measurable as well, and $\bar{\mu} g$ makes sense and it is clear that $\bar{\mu} g = \mu g$. Since f and g are $\bar{\mathcal{E}}$-measurable and $f = g$ $\bar{\mu}$-almost everywhere, $\bar{\mu} f = \bar{\mu} g$ by insensitivity. So, the convention above amounts to writing μf instead of $\bar{\mu} f$.

With this convention in place, we now re-state the monotone convergence theorem in full generality.

4.20 THEOREM. *Let (f_n) be a sequence of numerical functions on E. Suppose that, for each n, there is g_n in $\mathcal{E}$ such that $f_n = g_n$ almost everywhere. Further, suppose for each n that $f_n \ge 0$ almost everywhere and $f_n \le f_{n+1}$ almost everywhere. Then, $\lim f_n$ exists almost everywhere, is positive almost everywhere, and $\mu(\lim f_n) = \lim \mu f_n$.*

We discuss this fully to indicate its meaning and the issues involved. Let $\mathcal{N}$ denote the collection of all measurable negligible sets, that is, every N in $\mathcal{N}$ belongs to $\mathcal{E}$ and $\mu(N) = 0$. Now fix n. To say that $f_n = g_n$ almost everywhere is to say that there is N_n in $\mathcal{N}$ such that $f_n = g_n$ outside N_n (that is, $f_n(x) = g_n(x)$ whenever $x \notin N_n$). Similarly, $f_n \geq 0$ almost everywhere means that there is M_n in $\mathcal{N}$ such that $f_n \geq 0$ outside M_n. And, since $f_n \leq f_{n+1}$ almost everywhere, there is L_n in $\mathcal{N}$ such that $f_n \leq f_{n+1}$ outside L_n. These are the conditions. The claim of the theorem is as follows. First, there is an $\mathcal{E}$-measurable function f, and a set N in $\mathcal{N}$ such that $\lim f_n(x)$ exists and is equal to $f(x)$ for every x outside N. Also, there is M in $\mathcal{N}$ such that $f \geq 0$ outside M. Finally, $\mu f = \lim \mu f_n$, where the μf_n are defined by convention 4.19 to be the numbers μg_n.

Proof. Let

$$N = \bigcup_{n=1}^{\infty} (L_n \cup M_n \cup N_n).$$

Then, $N \in \mathcal{E}$ and $\mu(N) = 0$ by Boole's inequality, that is, $N \in \mathcal{N}$.

For x outside N, we have

$$0 \leq f_1(x) = g_1(x) \leq f_2(x) = g_2(x) \leq \ldots,$$

and hence $\lim f_n(x)$ exists and is equal to $\lim g_n(x)$. Define

$$f(x) = \begin{cases} \lim f_n(x) & \text{if } x \notin N \\ 0 & \text{if } x \in N \end{cases}$$

Clearly, f is the limit of the increasing sequence $(g_n 1_{E \setminus N})$ in $\mathcal{E}_+$. So, f is in $\mathcal{E}_+$ and we may take $M = \emptyset$. There remains to show that $\mu f = \lim \mu g_n$. Now in fact

$$\mu f = \mu(\lim g_n 1_{E \setminus N}) = \lim \mu(g_n 1_{E \setminus N}) = \lim \mu g_n,$$

where we used the monotone convergence theorem to justify the second equality, and the insensitivity to justify the third. □

The reader is invited to formulate the "almost everywhere version" of the dominated convergence theorem and to prove it carefully once. We shall use such versions without further ado whenever the need drives us.

Characterization of the integral

Definition 4.3 defines the integral μf for every f in $\mathcal{E}_+$. Thus, in effect, integration extends the domain of μ from the measurable sets (identified with their indicator functions) to the space $\mathcal{E}_+$ of all positive measurable functions (and beyond), and hence we may regard μ as the mapping $f \mapsto \mu f$ from $\mathcal{E}_+$ into $\bar{\mathbb{R}}_+$. The mapping $\mu : \mathcal{E}_+ \mapsto \bar{\mathbb{R}}_+$ is necessarily positive, linear, and continuous under increasing limits; these were promised in 4.2 and proved as Proposition 4.7, Proposition 4.12, and Theorem 4.8. We end this section with the following very useful converse.

4.21 THEOREM. *Let $(E, \mathcal{E})$ be a measurable space. Let L be a mapping from $\mathcal{E}_+$ into $\bar{\mathbb{R}}_+$. Then there exists a unique measure μ on $(E, \mathcal{E})$ such that $L(f) = \mu f$ for every f in $\mathcal{E}_+$ if and only if*

4.22 a) $f = 0 \Rightarrow L(f) = 0$.
b) $f, g \in \mathcal{E}_+$ and $a, b \in \mathbb{R}_+ \Rightarrow L(af + bg) = aL(f) + bL(g)$.
c) $(f_n) \subset \mathcal{E}_+$ and $f_n \nearrow f \Rightarrow L(f_n) \nearrow L(f)$.

Proof. Necessity of the conditions is immediate from the properties of the integral: (a) follows from the definition of μf, (b) from linearity, and (c) from the monotone convergence theorem.

To show the sufficiency, suppose that L has the properties (a)-(c). Define

4.23 $$\mu(A) = L(1_A), \qquad A \in \mathcal{E}.$$

We show that μ is a measure. First, $\mu(\emptyset) = L(1_\emptyset) = L(0) = 0$. Second, if $A_1, A_2, \ldots$ are disjoint sets in $\mathcal{E}$ with union A, then the indicator of $\bigcup_1^n A_i$ is $\sum_1^n 1_{A_i}$, the latter is increasing to 1_A, and hence,

$$\mu(A) = L(1_A) = \lim_n L(\sum_1^n 1_{A_i}) = \lim_n \sum_1^n L(1_{A_i}) = \lim_n \sum_1^n \mu(A_i) = \sum_1^\infty \mu(A_i),$$

where we used the conditions (c) and (b) to justify the second and third equality signs.

So, μ is a measure on $(E, \mathcal{E})$. It is unique by the necessity of 4.23. Now, $L(f) = \mu f$ for simple f in $\mathcal{E}_+$ by the linearity property (b) of L and the linearity of integration. This in turn implies that, for every f in $\mathcal{E}_+$, choosing simple $f_n \nearrow f$,

$$L(f) = \lim L(f_n) = \lim \mu f_n = \mu f$$

by condition (c) and the monotone convergence theorem. □

Exercises and complements

4.24 *Integrability.* If $f \in \mathcal{E}_+$ and $\mu f < \infty$, then f is real-valued almost everywhere. Show this. More generally, if f is integrable then it is real-valued almost everywhere.

4.25 *Test for vanishing.* Let $f \in \mathcal{E}_+$. Then $\mu f = 0$ if and only if $f = 0$ almost everywhere. Prove.

4.26 *Alternative form of the monotone convergence theorem.* If $f_1, f_2, \ldots$ are in $\mathcal{E}_+$ then

$$\mu \sum_1^\infty f_n = \sum_1^\infty \mu f_n.$$

4.27 *Sums of measures.* Recall that if $\mu_1, \mu_2, \ldots$ are measures on $(E, \mathcal{E})$, so is $\mu = \sum \mu_n$. Show that, for every f in $\mathcal{E}_+$,

$$\mu f = \sum_n \mu_n f.$$

4.28 *Absolute values.* Assuming that μf exists, show that $|\mu f| \le \mu |f|$.

4.29 *Mean value theorem.* If $\mu(A) > 0$ and $a \le f(x) \le b$ for every x in A, then show that

$$a \le \frac{1}{\mu(A)} \int_A f d\mu \ \le \ b.$$

4.30 *Generalization of the monotone convergence theorem.* If $f_n \ge g$ for all n for some integrable function g, and if (f_n) increases to f, then μf exists and is equal to $\lim \mu f_n$. If $f_n \le g$ for all n for some integrable function g and if (f_n) decreases to f, then μf exists and is equal to $\lim \mu f_n$.

4.31 *On dominated convergence.* In the dominated convergence theorem, the condition that (f_n) be dominated by an integrable g is necessary. Suppose that $E = (0,1)$, $\mathcal{E} = \mathcal{B}_E$, $\mu = Leb$. Take, for $n = 1, 2, \ldots$,

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Then, $f_n(x) \to 0$ for every x in E, the integral $\mu f_n = 1$ for every n, but $0 = \mu(\lim f_n) \ne \lim \mu f_n = 1$.

4.32 *Test for σ-finiteness.* A measure μ on $(E, \mathcal{E})$ is σ-finite if and only if there exists a strictly positive function f in $\mathcal{E}$ such that $\mu f < \infty$. Prove this. Hint for the sufficiency part: Let $E_n = \{f > \frac{1}{n}\}$ and note that $E_n \nearrow E$ whereas $\frac{1}{n}\mu(E_n) \le \mu(f 1_{E_n}) \le \mu f < \infty$.

5 Transforms and Indefinite Integrals

This section is about measures defined from other measures via various means and the relationships among integrals with respect to them.

Image measures

Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. Let ν be a measure on $(F, \mathcal{F})$ and let $h : F \mapsto E$ be measurable relative to $\mathcal{F}$ and $\mathcal{E}$. We define a mapping $\nu \circ h^{-1}$ from the σ-algebra $\mathcal{E}$ into $\bar{\mathbb{R}}_+$ by

5.1 $$\nu \circ h^{-1}(B) = \nu(h^{-1}B), \qquad B \in \mathcal{E},$$

which is well-defined since $h^{-1}B \in \mathcal{F}$ by the measurability of h. It is easy to check that $\nu \circ h^{-1}$ is a measure on $(E, \mathcal{E})$; it is called the *image* of ν under h. Other notations current are $h \circ \nu$, $h(\nu)$, $\nu \circ h$, ν_h.

If ν is finite, then so is its image. If ν is Σ-finite, again, so is its image. But, the image of a σ-finite measure generally fails to be σ-finite (but is Σ-finite).

The following relates integrals with respect to $\nu \circ h^{-1}$ to integrals with respect to ν.

5.2 THEOREM. *For every f in $\mathcal{E}_+$ we have $(\nu \circ h^{-1})f = \nu(f \circ h)$.*

Proof. Define $L : \mathcal{E}_+ \mapsto \bar{\mathbb{R}}_+$ by setting $L(f) = \nu(f \circ h)$. It can be checked that L satisfies the conditions of the integral characterization theorem 4.21. Thus, $L(f) = \mu f$ for some unique measure μ on $(E, \mathcal{E})$. That μ is precisely the measure $\nu \circ h^{-1}$, because

$$\mu(B) = L(1_B) = \nu(1_B \circ h) = \nu(h^{-1}B), \qquad B \in \mathcal{E}. \qquad \square$$

The limitation to positive $\mathcal{E}$-measurable functions can be removed: for arbitrary f in $\mathcal{E}$ the same formula holds provided that the integral on one side be well-defined (and then both sides are well-defined).

The preceding theorem is a generalization of the change of variable formula from calculus. In more explicit notation, with $\mu = \nu \circ h^{-1}$, the theorem is that

5.3 $$\int_F \nu(dx) f(h(x)) = \int_E \mu(dy) f(y),$$

that is, if $h(x)$ is replaced with y then $\nu(dx)$ must be replaced with $\mu(dy)$. In calculus, it is often the case that $E = F = \mathbb{R}^d$ for some fixed dimension d, and μ and ν are expressed in terms of the Lebesgue measure on $\mathbb{R}^d$ and the Jacobian of the transformation h. In probability theory, often, the measure ν is defined implicitly through the formula 5.3 by stating the transformation h and the corresponding image measure μ. We take up still another use next.

Images of the Lebesgue measure

Forming image measures is a convenient method of creating new measures from the old, and if the old measure ν is convenient enough as an integrator, then 5.3 provides a useful formula for the integrals with respect to the new measure μ. In fact, the class of measures that can be represented as images of the Lebesgue measure on $\mathbb{R}_+$ is very large. The following is the precise statement; combined with the preceding theorem it reduces integrals over abstract spaces to integrals on $\mathbb{R}_+$ with respect to the Lebesgue measure.

5.4 THEOREM. *Let $(E, \mathcal{E})$ be a standard measurable space. Let μ be a Σ-finite measure on $(E, \mathcal{E})$ and put $b = \mu(E)$, possibly $+\infty$. Then, there exists a mapping h from $[0, b)$ into E, measurable relative to $\mathcal{B}_{[0,b)}$ and $\mathcal{E}$, such that*

$$\mu = \lambda \circ h^{-1},$$

where λ is the Lebesgue measure on $[0, b)$.

Proof will be sketched in Exercises 5.15 and 5.16 in a constructive fashion.

Indefinite integrals

Let $(E, \mathcal{E}, \mu)$ be a measure space. Let p be a positive $\mathcal{E}$-measurable function. Define

5.5 $$\nu(A) = \mu(p1_A) = \int_A \mu(dx)p(x), \qquad A \in \mathcal{E}.$$

It follows from the monotone convergence theorem (alternative form) that ν is a measure on $(E, \mathcal{E})$. It is called the *indefinite integral* of p with respect to μ.

5.6 PROPOSITION. *For every f in $\mathcal{E}_+$, we have $\nu f = \mu(pf)$.*

Proof. Let $L(f) = \mu(pf)$ and check that L satisfies the conditions of Theorem 4.21. Thus, there exists a unique measure $\hat{\mu}$ on $(E, \mathcal{E})$ such that $L(f) = \hat{\mu} f$ for every f in $\mathcal{E}_+$. We have $\hat{\mu} = \nu$, since

$$\hat{\mu}(A) = L(1_A) = \mu(p1_A) = \nu(A), \qquad A \in \mathcal{E}. \qquad \square$$

The formula 5.5 is another convenient tool for creating new measures from the old. Written in more explicit notation, the preceding proposition becomes

5.7 $$\int_E \nu(dx)\, f(x) = \int_E \mu(dx)\, p(x)\, f(x) \qquad f \in \mathcal{E}_+,$$

which can be expressed informally by writing

5.8 $$\nu(dx) = \mu(dx)\, p(x), \qquad x \in E,$$

once it is understood that μ and ν are measures on $(E, \mathcal{E})$ and that p is positive $\mathcal{E}$-measurable.

Heuristically, we may think of $\mu(dx)$ as the amount of mass put by μ on an "infinitesimal neighborhood" dx of the point x, and similarly of $\nu(dx)$. Then, 5.8 takes on the meaning that $p(x)$ is the mass density, at x, of the measure ν with respect to μ. For this reason, the function p is called the density function of ν relative to μ, and the following notations are used for it:

5.9 $$p = \frac{d\nu}{d\mu}; \qquad p(x) = \frac{\nu(dx)}{\mu(dx)}, \qquad x \in E.$$

The expressions 5.5-5.9 are equivalent ways of saying the same thing: ν is the indefinite integral of p with respect to μ, or p is the density of ν relative to μ.

Radon-Nikodym theorem

Let μ and ν be measures on a measurable space $(E, \mathcal{E})$. Then, ν is said to be *absolutely continuous* with respect to μ if, for every set A in $\mathcal{E}$,

5.10 $$\mu(A) = 0 \Rightarrow \nu(A) = 0.$$

If ν is the indefinite integral of some positive $\mathcal{E}$-measurable function with respect to μ, then it is evident from 5.5 that ν is absolutely continuous with respect to μ. The following, called the Radon-Nikodym theorem, shows that the converse is true as well, at least when μ is σ-finite. We list it here without proof. We shall give two proofs of it later.

5.11 THEOREM. *Suppose that μ is σ-finite, and ν is absolutely continuous with respect to μ. Then, there exists a positive $\mathcal{E}$-measurable function p such that*

5.12
$$\int_E \nu(dx)\, f(x) = \int_E \mu(dx)\, p(x)\, f(x), \qquad f \in \mathcal{E}_+.$$

Moreover, p is unique up to equivalence: if 5.12 holds for another $\hat{p}$ in $\mathcal{E}_+$, then $\hat{p}(x) = p(x)$ for μ-almost every x in E.

The function p in question can be denoted by $d\nu/d\mu$ in view of the equivalence of 5.5-5.9 and 5.12; and the function p is also called the *Radon-Nikodym derivative* of ν with respect to μ. See Exercises 5.17-5.20 for some remarks.

A matter of style

When an explicit expression is desired for a measure μ, there are several choices. One can go with the definition and give a formula for $\mu(A)$. Equivalently, and usually with greater ease and clarity, one can display a formula for the integral μf for arbitrary f in $\mathcal{E}_+$. In those cases where μ has a density with respect to some well-known measure like the Lebesgue measure, it is better to give the formula for μf or, to be more brief, to give a formula like $\mu(dx) = \lambda(dx)\, p(x)$ by using the form 5.8, with λ denoting the Lebesgue measure. All things considered, if a uniform style is desired, it is best to display an expression for μf. We shall do either that or use the form 5.8 when the form of p is important.

Exercises and complements

5.13 *Time changes.* Let c be an increasing right-continuous function from $\mathbb{R}_+$ into $\bar{\mathbb{R}}_+$. Define

$$a(u) = \inf\{t \in \mathbb{R}_+ :\ c(t) > u\}, \qquad u \in \mathbb{R}_+,$$

with the usual convention that $\inf \emptyset = \infty$.

a) Show that the function $a : \mathbb{R}_+ \mapsto \bar{\mathbb{R}}_+$ is increasing and right-continuous, and that

$$c(t) = \inf\{u \in \mathbb{R}_+ :\ a(u) > t\}, \qquad t \in \mathbb{R}_+.$$

Thus, a and c are right-continuous "functional inverses" of each other. See Figure 1 below.

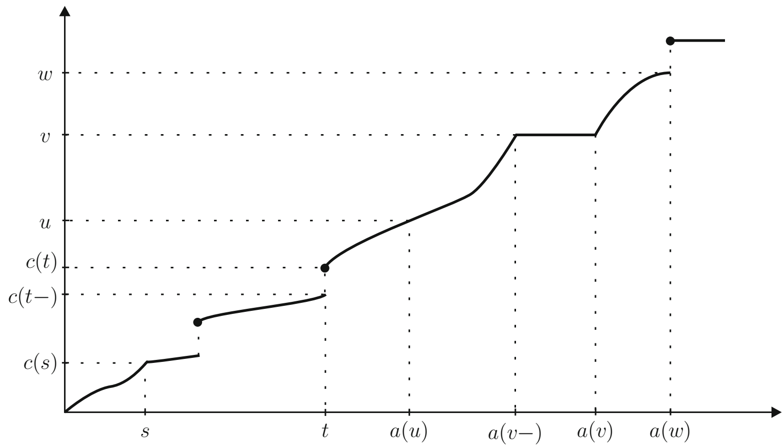


Figure 1: Both c and a are increasing right-continuous. They are functional inverses of each other.

b) Suppose $c(t) < \infty$. Show that $a(c(t)) \geq t$, with equality if and only if $c(t+\varepsilon) > c(t)$ for every $\varepsilon > 0$.

Imagine a clock whose mechanism is so rigged that it points to the number $c(t)$ when the actual time is t. Then, when the clock points to the number u, the actual time is $a(u)$. Hence the term “time change” for the operations involved.

5.14 *Distribution functions and measures on* $\mathbb{R}_+$. Let μ be a measure on $\mathbb{R}_+$ (with its Borel σ-algebra) such that $c(t) = \mu[0,t]$ is finite for every t in $\mathbb{R}_+$. The limit $b = c(\infty) = \lim_{t\to\infty} c(t)$ is allowed to be $+\infty$.

a) Show that c is increasing and right-continuous. It is called the *cumulative distribution function* associated with μ.

b) Define $a(u)$ as in 5.13 for $u \in [0,b)$, let λ denote the Lebesgue measure on $[0,b)$. Show that

$$\mu = \lambda \circ a^{-1} .$$

This demonstrates Theorem 5.4 in the case of measures like the present μ. Incidentally, we have also shown that to every increasing right-continuous function c from $\mathbb{R}_+$ into $\mathbb{R}_+$ there corresponds a unique measure μ on $\mathbb{R}_+$ whose cumulative distribution function is c.

5.15 *Representation of measures: Finite case.* Let μ be a finite measure on a standard measurable space $(E, \mathcal{E})$. We aim to prove Theorem 5.4 in this case assuming that $(E, \mathcal{E})$ is isomorphic to $(D, \mathcal{B}_D)$ where $D = [0,1]$. The remaining cases where E is finite or countably infinite are nearly trivial.

The idea is simple: First, use the isomorphism to carry μ from E into a measure $\hat{\mu}$ on D. Second, follow the steps of 5.14 to write $\hat{\mu} = \lambda \circ a^{-1}$ where λ is the Lebesgue measure on $B = [0, b)$ with $b = \hat{\mu}(D) = \mu(E)$. Finally, use the inverse of the isomorphism to carry $\hat{\mu}$ back onto E. Here are the details.
Let $f: E \mapsto D$ be the isomorphism involved. Let $g: D \mapsto E$ be the functional inverse of f, that is, $g(t) = x$ if and only if $f(x) = t$. Define $\hat{\mu} = \mu \circ f^{-1}$; then $\hat{\mu}$ is a measure on D with total mass $\hat{\mu}(D) = \mu(E) = b$. Put $B = [0, b)$, $\mathcal{B} = \mathcal{B}_B$ and λ the Lebesgue measure on B.
Define $c(t) = \hat{\mu}[0, t]$ for t in D. Define $a(u)$ by 5.13 for u in B. Note that $a: B \mapsto D$ is measurable and that $\hat{\mu} = \lambda \circ a^{-1}$. Define $h(u) = g \circ a(u)$ for u in B. Observe that $\lambda \circ h^{-1} = \lambda \circ a^{-1} \circ g^{-1} = \mu$ as needed.

5.16 *Continuation: Σ-finite case.* Let $(E, \mathcal{E})$ be isomorphic to $(D, \mathcal{B}_D)$ where $D = [0, 1]$. Let μ be Σ-finite on $(E, \mathcal{E})$, say $\mu = \sum \mu_n$ with each μ_n finite. Since the case of finite μ is already covered, we assume that $b = \mu(E) = +\infty$. Let $B = [0, b) = \mathbb{R}_+$, $\mathcal{B} = \mathcal{B}(\mathbb{R}_+)$, and λ the Lebesgue measure on $\mathbb{R}_+$. Let $f: E \mapsto D$ and $g: D \mapsto E$ as before.
Let $D_n = 2n + D = [2n, 2n + 1]$, $n = 0, 1, 2, \ldots$; note that $D_0, D_1, \ldots$ are disjoint. Define $f_n: E \mapsto D_n$ by setting $f_n(x) = 2n + f(x)$ and let $g_n: D_n \mapsto E$ be the functional inverse of f_n, that is, $g_n(t) = g(t - 2n)$. Now, $\hat{\mu}_n = \mu_n \circ f_n^{-1}$ is a measure on D_n. Define

$$\hat{\mu}(C) = \sum_0^{\infty} \hat{\mu}_n(C \cap D_n), \qquad C \in \mathcal{B}(\mathbb{R}_+).$$

This defines a measure $\hat{\mu}$ on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $c(t) = \hat{\mu}[0, t] < \infty$ for every t in $\mathbb{R}_+$, as in Exercise 5.14. Let a be defined as in 5.13, and observe that $\hat{\mu} = \lambda \circ a^{-1}$. Also observe that, by the way a is defined, $a(u)$ belongs to the set $\bigcup_n D_n$ for each u. Finally, put

$$h(u) = g_n \circ a(u) \qquad \text{if } a(u) \in D_n,$$

and show that $\mu = \lambda \circ h^{-1}$ as claimed.

5.17 *Absolute continuity for atomic measures.* Let ν be a Σ-finite purely atomic measure on some measurable space $(E, \mathcal{E})$ such that the singletons $\{x\}$ belong to $\mathcal{E}$ for each x in E. Let D be the collection of all atoms, and recall that D is countable. Let $\mu(A)$ be the number of points in $A \cap D$. Then, ν is absolutely continuous with respect to μ. Find the density $p = d\nu/d\mu$.

5.18 *Radon-Nikodym derivatives.* Let μ be a measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $c(t) = \mu[0, t]$ is finite for every t in $\mathbb{R}_+$. If μ is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}_+$, then the cumulative distribution function c is differentiable at λ-almost every t in $\mathbb{R}_+$ and

$$p(t) = \frac{\mu(dt)}{\lambda(dt)} = \frac{d}{dt} c(t) \qquad \text{for } \lambda\text{−almost every } t.$$

5.19 *Radon-Nikodym and σ-finiteness.* The condition that μ be σ-finite cannot be removed in general. Let ν be the Lebesgue measure on $E = [0,1]$, $\mathcal{E} = \mathcal{B}(E)$, and let $\mu(A)$ be 0 or $+\infty$ according as $\nu(A)$ is 0 or strictly positive, $A \in \mathcal{E}$. Then, μ is Σ-finite, and ν is absolutely continuous with respect to μ. Show that the conclusion of Theorem 5.11 fails in this case.

5.20 *On Σ-finiteness.* Let μ be a Σ-finite measure on an arbitrary measurable space $(E, \mathcal{E})$, say with the decomposition $\mu = \sum \mu_n$, where $\mu_n(E) < \infty$ for each n. Define $\nu(A) = \sum_n \mu_n(A)/2^n \mu_n(E)$, $A \in \mathcal{E}$. Show that ν is a finite measure, and μ is absolutely continuous with respect to ν. Thus, there exists $p \in \mathcal{E}_+$ such that

$$\mu(dx) = \nu(dx)\, p(x)\,, \qquad x \in E.$$

If μ is σ-finite, show that p is real-valued ν-almost everywhere. Show that, in the converse direction, if μ is absolutely continuous with respect to a finite measure ν, then μ is Σ-finite.

5.21 *Singularity.* Let μ and ν be measures on some measurable space $(E, \mathcal{E})$. Then, ν is said to be *singular* with respect to μ if there exists a set D in $\mathcal{E}$ such that

$$\mu(D) = 0 \qquad \text{and} \qquad \nu(E \setminus D) = 0.$$

The notion is the opposite of absolute continuity. Show that, if ν is purely atomic and μ is diffuse then ν is singular with respect to μ. This does not exhaust the possibilities, however, as the famous example next illustrates.

5.22 *Cantor set, Cantor measure.* Start with the interval $E = [0,1]$. Delete the set $D_{0,1} = (\frac{1}{3}, \frac{2}{3})$ which forms the middle third of E; this leaves two closed intervals. Delete the middle thirds of those, that is, delete $D_{1,1} = (\frac{1}{9}, \frac{2}{9})$ and $D_{1,2} = (\frac{7}{9}, \frac{8}{9})$; there remain four closed intervals. Delete the middle thirds of those four intervals, and continue in this fashion. At the end, the deleted intervals form the open set

$$D = \bigcup_{i=0}^{\infty} \bigcup_{j=1}^{2^i} D_{i,j},$$

and the set of points that remain is

$$C = E \setminus D.$$

The closed set C is called the *Cantor set.*

Next we construct a continuous function $c: E \mapsto [0,1]$ that remains constant over each interval $D_{i,j}$ and increases (only) on C. Define $c(t) = \frac{1}{2}$ for t in $D_{0,1}$; let $c(t) = \frac{1}{4}$ for t in $D_{1,1}$ and $c(t) = \frac{3}{4}$ for t in $D_{1,2}$; let $c(t) = \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$ according as t is in $D_{2,1}$, $D_{2,2}$, $D_{2,3}$, $D_{2,4}$; and so on. This defines a uniformly continuous increasing function from D into $[0,1]$. Since D is dense in E, we may extend c onto E by continuity. The resulting function $c: E \mapsto [0,1]$ is called the Cantor function.

a) Show that $\text{Leb}(D) = 1$, $\text{Leb}(C) = 0$.

b) Let ν be the measure on E corresponding to the (cumulative distribution) function c, that is, $\nu = \lambda \circ a^{-1}$ where λ is the Lebesgue measure on $[0, 1)$ and $a : [0, 1) \mapsto E$ is the inverse of c as in 5.13. We call ν the Cantor measure. Show that $\nu(C) = 1$ and $\nu(D) = 0$. Conclude that ν is a diffuse measure on E and that ν is singular with respect to the Lebesgue measure on E.

c) Show that the range of a is $C \setminus C_0$ where C_0 consists of the point 1 and the countable collection of points that are the left-end-points of the intervals $D_{i,j}$. Thus, a is a one-to-one mapping from $[0, 1)$ onto $C \setminus C_0$, and it follows that $C \setminus C_0$ has the power of the continuum. Thus, the Cantor set has the power of the continuum, even though its Lebesgue measure is 0.

d) The Cantor set is everywhere dense in itself, that is, for every t in C there exists $(t_n) \subset C\setminus\{t\}$ such that $t = \lim t_n$. Incidentally, a closed set that is everywhere dense in itself is said to be *perfect.*

5.23 *Lebesgue-Stieltjes integrals.* Let c be an increasing right-continuous function from $\mathbb{R}_+$ into $\mathbb{R}_+$. Let μ be the measure on $\mathbb{R}_+$ that has c as its cumulative distribution function (see Exercise 5.14). For each positive Borel function f on $\mathbb{R}_+$, define

$$\int_{\mathbb{R}_+} f(t)\, dc(t) = \int_{\mathbb{R}_+} \mu(dt)\, f(t).$$

The left side is called the Lebesgue-Stieltjes integral of f with respect to c. Note that, with the notation of 5.14,

$$\int_{\mathbb{R}_+} \mu(dt) f(t) = \int_0^b du f(a(u)).$$

Replacing f by $f1_A$ one obtains the same integral over the interval A. Extensions to arbitrary Borel functions f on $\mathbb{R}_+$ are as usual for μf, namely, by using the decomposition $f = f^+ - f^-$. Extension from the space $\mathbb{R}_+$ onto $\mathbb{R}$ is obvious. Finally, extensions to functions c that can be decomposed as $c = c_1 - c_2$ with both c_1 and c_2 increasing and right-continuous (see the next exercise) can be done by setting

$$\int_{\mathbb{R}} f(t)\, dc(t) = \int_{\mathbb{R}} f(t)\, dc_1(t) - \int_{\mathbb{R}} f(t)\, dc_2(t)$$

for those f for which the integrals on the right make sense and are not both $+\infty$ or both $-\infty$.

5.24 *Functions of bounded variation.* Let f be a function from $\mathbb{R}_+$ into $\mathbb{R}$. Think of $f(t)$ as the position, at time t, of an insect moving on the line $\mathbb{R}$. We are interested in the total amount of traveling done during a finite interval $(s, t]$. Here is the precise version.

A *subdivision* of $[s,t]$ is a finite collection $\mathcal{A}$ of disjoint intervals of the form $(\ ,\ ]$ whose union is $(s,t]$. We define

$$V_f(s,t) = \sup_{\mathcal{A}} \sum_{(u,v]\in\mathcal{A}} |f(v) - f(u)|$$

where the supremum is over all subdivisions $\mathcal{A}$ of $[s,t]$. The number $V_f(s,t)$ is called the *total variation* of f on $(s,t]$. The function f is said to be of *bounded variation* on $[s,t]$ if $V_f(s,t) < \infty$.

Show the following:

a) If f is increasing on $[s,t]$, then $V_f(s,t) = f(t) - f(s)$.

b) If f is differentiable and its derivative is bounded by b on $[s,t]$, then $V_f(s,t) \le (t-s)\cdot b$.

c) $V_f(s,t) + V_f(t,u) = V_f(s,u)$ for $s < t < u$.

d) $V_{f+g}(s,t) \le V_f(s,t) + V_g(s,t)$. Thus, if f and g are of bounded variation on $[s,t]$, then so are $f+g$ and $f-g$.

e) The function f is of bounded variation on $[s,t]$ if and only if $f = g - h$ for some real-valued positive functions g and h that are increasing on $[s,t]$.

Hint: To show the necessity, define $g(r)$ and $h(r)$ for r in $(s,t]$ by

$$2g(r) = V_f(s,r) + f(r) + f(s), \qquad 2h(r) = V_f(s,r) - f(r) + f(s)$$

and show that g and h are increasing and $f = g - h$.

The class of functions f for which Lebesgue-Stieltjes integrals $\int g\,df$ are defined is the class of f that are of bounded variation over bounded intervals.

6 Kernels and Product Spaces

Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces. Let K be a mapping from $E\times\mathcal{F}$ into $\bar{\mathbb{R}}_+$. Then, K is called a *transition kernel* from $(E,\mathcal{E})$ into $(F,\mathcal{F})$ if

6.1 a) the mapping $x \mapsto K(x,B)$ is $\mathcal{E}$-measurable for every set B in $\mathcal{F}$, and

b) the mapping $B \mapsto K(x,B)$ is a measure on $(F,\mathcal{F})$ for every x in E.

For example, if ν is a finite measure on $(F,\mathcal{F})$, and k is a positive function on $E\times F$ that is measurable with respect to the product σ-algebra $\mathcal{E}\otimes\mathcal{F}$, then it will be seen shortly that

6.2 $$K(x,B) = \int_B \nu(dy)\,k(x,y)\,, \qquad x\in E,\ B\in\mathcal{F}\,,$$

defines a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. In the further special case where $E = \{1, \ldots, m\}$ and $F = \{1, \ldots, n\}$ with their discrete σ-algebras, the transition kernel K is specified by the numbers $K(x, \{y\})$ and can be regarded as an m by n matrix of positive numbers. This special case will inform the choice of notations like Kf and μK below (recall that functions are thought as generalizations of column vectors and measures as generalizations of row vectors).

Measure-kernel-function

6.3 THEOREM. *Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Then,*

$$Kf(x) = \int_F K(x, dy)\, f(y)\,, \qquad x \in E,$$

defines a function Kf that is in $\mathcal{E}_+$ for every function f in $\mathcal{F}_+$;

$$\mu K(B) = \int_E \mu(dx)\, K(x, B)\,, \qquad B \in \mathcal{F},$$

defines a measure μK on $(F, \mathcal{F})$ for each measure μ on $(E, \mathcal{E})$; and

$$(\mu K)f = \mu(Kf) = \int_E \mu(dx) \int_F K(x, dy)\, f(y)$$

for every measure μ on $(E, \mathcal{E})$ and function f in $\mathcal{F}_+$.

Proof. a) Let $f \in \mathcal{F}_+$. Then Kf is a well-defined positive function on E, since the number $Kf(x)$ is the integral of f with respect to the measure $B \mapsto K(x, B)$. We show that Kf is $\mathcal{E}$-measurable in two steps: First, if f is simple, say $f = \sum_1^n b_i 1_{B_i}$, then $Kf(x) = \sum b_i K(x, B_i)$, which shows that Kf is $\mathcal{E}$-measurable since it is a linear combination of the $\mathcal{E}$-measurable functions $x \mapsto K(x, B_i)$, $i = 1, \ldots n$. Second, if f in $\mathcal{F}_+$ is not simple, we choose simple f_n in $\mathcal{F}_+$ increasing to f; then $Kf(x) = \lim_n Kf_n(x)$ for each x by the monotone convergence theorem for the measure $B \mapsto K(x, B)$; and, hence Kf is $\mathcal{E}$-measurable since it is the limit of $\mathcal{E}$-measurable functions Kf_n.

b) We prove the remaining two claims together. Fix a measure μ on $(E, \mathcal{E})$. Define $L:\ \mathcal{F}_+ \mapsto \bar{\mathbb{R}}_+$ by setting

$$L(f) = \mu(Kf).$$

If $f = 0$ then $L(f) = 0$. If f and g are in $\mathcal{F}_+$, and a and b in $\mathbb{R}_+$, then

$$\begin{aligned} L(af + bg) &= \mu(K(af + bg)) = \mu(aKf + bKg) \\ &= a\mu(Kf) + b\mu(Kg) = aL(f) + bL(g), \end{aligned}$$

where the second equality is justified by the linearity of the integration with respect to the measure $B \mapsto K(x, B)$ for each x, and the third equality by

the linearity of the integration with respect to μ. Finally, if $(f_n) \subset \mathcal{F}_+$ and $f_n \nearrow f$, then $Kf_n(x) \nearrow Kf(x)$ by the monotone convergence theorem for $B \mapsto K(x, B)$, and

$$L(f_n) = \mu(Kf_n) \nearrow \mu(Kf) = L(f)$$

by the monotone convergence theorem for μ. Hence, by Theorem 4.21, there exists a measure ν on $(F, \mathcal{F})$ such that $L(f) = \nu f$ for every f in $\mathcal{F}_+$. Taking $f = 1_B$, we see that $\nu(B) = \mu K(B)$ for every set B in $\mathcal{F}$, that is, $\nu = \mu K$. So, μK is a measure on $(F, \mathcal{F})$, and $(\mu K)f = \nu f = L(f) = \mu(Kf)$ as claimed.

6.4 REMARK. To specify a kernel K from $(E, \mathcal{E})$ into $(F, \mathcal{F})$ it is more than enough to specify Kf for every f in $\mathcal{F}_+$. Conversely, as an extension of Theorem 4.21, it is easy to see that a mapping $f \mapsto Kf$ from $\mathcal{F}_+$ into $\mathcal{E}_+$ specifies a transition kernel K if and only if

a) $K0 = 0$,
b) $K(af + bg) = aKf + bKg$ for f and g in $\mathcal{F}_+$ and a and b in $\mathbb{R}_+$,
c) $Kf_n \nearrow Kf$ for every sequence (f_n) in $\mathcal{F}_+$ increasing to f.

Obviously, then, $K(x, B) = K1_B(x)$.

Products of kernels, Markov kernels

Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$ and let L be a transition kernel from $(F, \mathcal{F})$ into $(G, \mathcal{G})$. Then, their *product* is the transition kernel KL from $(E, \mathcal{E})$ into $(G, \mathcal{G})$ defined by

6.5 $$(KL)f = K(Lf), \qquad f \in \mathcal{G}_+.$$

Remark 6.4 above can be used to show that KL is indeed a kernel. Obviously,

$$KL(x, B) = \int_F K(x, dy)\, L(y, B)\,, \qquad x \in E,\ B \in \mathcal{G}.$$

A transition kernel from $(E, \mathcal{E})$ into $(E, \mathcal{E})$ is called simply a transition kernel on $(E, \mathcal{E})$. Such a kernel K is called a *Markov* kernel on $(E, \mathcal{E})$ if $K(x, E) = 1$ for every x, and a *sub-Markov* kernel if $K(x, E) \le 1$ for every x.

If K is a transition kernel on $(E, \mathcal{E})$, its *powers* are the kernels on $(E, \mathcal{E})$ defined recursively by

6.6 $$K^0 = I, \quad K^1 = K, \quad K^2 = KK, \quad K^3 = KK^2, \ldots,$$

where I is the identity kernel on $(E, \mathcal{E})$:

6.7 $$I(x, A) = \delta_x(A) = 1_A(x), \qquad x \in E,\ A \in \mathcal{E}.$$

Note that $If = f$, $\mu I = \mu$, $\mu If = \mu f$, $IK = KI = K$ always. If K is Markov, so is K^n for every integer $n \ge 0$.

Kernels finite and bounded

Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. In analogy with measures, K is said to be *finite* if $K(x, F) < \infty$ for each x, and σ-*finite* if $B \mapsto K(x, B)$ is σ-finite for each x. It is said to be *bounded* if $x \mapsto K(x, F)$ is bounded, and σ-*bounded* if there exists a measurable partition (F_n) of F such that $x \mapsto K(x, F_n)$ is bounded for each n. It is said to be Σ-*finite* if $K = \sum_1^\infty K_n$ for some sequence of finite kernels K_n, and Σ-*bounded* if the K_n can be chosen to be bounded. In the very special case where $K(x, F) = 1$ for all x, the kernel is said to be a *transition probability* kernel. Markov kernels are transition probability kernels. Some connections between these notions are put in exercises.

Functions on product spaces

We start by re-stating the content of Exercise 2.22: sections of a measurable function are measurable.

6.8 Proposition. *Let $f \in \mathcal{E} \otimes \mathcal{F}$. Then, $x \mapsto f(x, y)$ is in $\mathcal{E}$ for each y in F, and $y \mapsto f(x, y)$ is in $\mathcal{F}$ for each x in E.*

Unfortunately, the converse is not true: it is possible that the conclusions hold, and yet f is not $\mathcal{E} \otimes \mathcal{F}$-measurable. One needs something stronger than measurability in at least one of the variables to conclude that f is in $\mathcal{E} \otimes \mathcal{F}$. See Exercise 6.28 for such an example.

The following is a generalization of the operation $f \mapsto Kf$ of Theorem 6.3 to functions f defined on the product space.

6.9 Proposition. *Let K be a Σ-finite kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Then, for every positive function f in $\mathcal{E} \otimes \mathcal{F}$,*

6.10 $$Tf(x) = \int_F K(x, dy)\, f(x, y), \qquad x \in E,$$

defines a function Tf in $\mathcal{E}_+$. Moreover, the transformation $T : (\mathcal{E} \otimes \mathcal{F})_+ \mapsto \mathcal{E}_+$ is linear and continuous under increasing limits, that is,

a) $T(af + bg) = aTf + bTg$ *for positive f and g in $\mathcal{E} \otimes \mathcal{F}$, and a and b in $\mathbb{R}_+$,*

b) $Tf_n \nearrow Tf$ *for every positive sequence $(f_n) \subset \mathcal{E} \otimes \mathcal{F}$ with $f_n \nearrow f$.*

Proof. Let f be a positive function in $\mathcal{E} \otimes \mathcal{F}$. Then, for each x in E, the section $f_x : \ y \mapsto f(x, y)$ is a function in $\mathcal{F}_+$ by Proposition 6.8, and $Tf(x)$ is the integral of f_x with respect to the measure $K_x : \ B \mapsto K(x, B)$. Thus, $Tf(x)$ is a well-defined positive number for each x, and the linearity

property (a) is immediate from the linearity of integration with respect to K_x for all x, and the continuity property (b) follows from the monotone convergence theorem for the measures K_x. There remains to show that Tf is $\mathcal{E}$-measurable.

We show this by a monotone class argument assuming that K is bounded. Boundedness of K implies that Tf is well-defined by 6.10 and is bounded for each bounded f in $\mathcal{E} \otimes \mathcal{F}$, and it is checked easily that

$$\mathcal{M} = \{f \in \mathcal{E} \otimes \mathcal{F} : \quad f \text{ is positive or bounded, } Tf \in \mathcal{E}\}$$

is a monotone class. Moreover, $\mathcal{M}$ includes the indicator of every measurable rectangle $A \times B$, since

$$T1_{A\times B}(x) = \int_F K(x, dy) 1_A(x) 1_B(y) = 1_A(x) K(x, B)$$

and the right side defines an $\mathcal{E}$-measurable function. Since the measurable rectangles generate the σ-algebra $\mathcal{E} \otimes \mathcal{F}$, it follows from the monotone class theorem 2.19 that $\mathcal{M}$ includes all positive (or bounded) f in $\mathcal{E} \otimes \mathcal{F}$ assuming that K is bounded. See Exercise 6.29 for extending the proof to Σ-finite K. □

Measures on the product space

The following is the general method for constructing measures on the product space $(E \times F, \mathcal{E} \otimes \mathcal{F})$.

6.11 THEOREM. *Let μ be a measure on $(E, \mathcal{E})$. Let K be a Σ-finite transition kernel from $(E, \mathcal{E})$ to $(F, \mathcal{F})$. Then,*

6.12 $$\pi f = \int_E \mu(dx) \int_F K(x, dy) f(x, y), \qquad f \in (\mathcal{E} \otimes \mathcal{F})_+$$

defines a measure π on the product space $(E \times F, \mathcal{E} \otimes \mathcal{F})$. Moreover, if μ is σ-finite and K is σ-bounded, then π is σ-finite and is the unique measure on that product space satisfying

6.13 $$\pi(A \times B) = \int_A \mu(dx) K(x, B), \qquad A \in \mathcal{E},\ B \in \mathcal{F}.$$

Proof. In the notation of the last proposition, the right side of 6.12 is $\mu(Tf)$, the integral of Tf with respect to μ. To see that it defines a measure, we use Theorem 4.21. Define $L(f) = \mu(Tf)$ for f in $\mathcal{E} \otimes \mathcal{F}$ positive. Then, $L(0) = 0$ obviously, L is linear since T is linear and integration is linear, and L is continuous under increasing limits by the same property for T and the monotone convergence theorem for μ. Hence, there is a unique measure, call it π, such that $L(f)$ is the integral of f with respect to π for every positive f in $\mathcal{E} \otimes \mathcal{F}$. This proves the first claim.

To prove the second, start by observing that π satisfies 6.13. Supposing that μ is σ-finite and K is σ-bounded, there remains to show that π is σ-finite and is the only measure satisfying 6.13. To that end, let $\hat{\pi}$ be another measure satisfying 6.13. Since μ is σ-finite, there is a measurable partition (E_m) of E such that $\mu(E_m) < \infty$ for each m. Since K is σ-bounded, there is a measurable partition (F_n) of F such that $x \mapsto K(x, F_n)$ is bounded for each n. Note that the measurable rectangles $E_m \times F_n$ form a partition of $E \times F$ and that, by the formula 6.13 for π and $\hat{\pi}$,

$$\pi(E_m \times F_n) = \hat{\pi}(E_m \times F_n) < \infty$$

for each m and n. Thus, the measures π and $\hat{\pi}$ are σ-finite, they agree on the p-system of measurable rectangles generating $\mathcal{E} \otimes \mathcal{F}$, and that p-system contains a partition of $E \times F$ over which π and $\hat{\pi}$ are finite. It follows from Exercise 3.18 that $\pi = \hat{\pi}$. □

Product measures and Fubini

In the preceding theorem, if the kernel K has the special form $K(x, B) = \nu(B)$ for some Σ-finite measure ν on $(F, \mathcal{F})$, then the measure π is called the *product* of μ and ν and is denoted by $\mu \times \nu$. The following theorem, generally referred to as Fubini's, is concerned with integration with respect to $\pi = \mu \times \nu$. Its main point is the formula 6.15: under reasonable conditions, in repeated integration, one can change the order of integration with impunity.

6.14 THEOREM. *Let μ and ν be Σ-finite measures on $(E, \mathcal{E})$ and $(F, \mathcal{F})$, respectively.*

a) There exists a unique Σ-finite measure π on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ such that, for every positive f in $\mathcal{E} \otimes \mathcal{F}$,

6.15 $$\pi f = \int_E \mu(dx) \int_F \nu(dy)\, f(x, y) = \int_F \nu(dy) \int_E \mu(dx)\, f(x, y).$$

b) If $f \in \mathcal{E} \otimes \mathcal{F}$ and is π-integrable, then $y \mapsto f(x, y)$ is ν-integrable for μ-almost every x, and $x \mapsto f(x, y)$ is μ-integrable for ν-almost every y, and 6.15 holds again.

6.16 REMARK. a) Since we have more than one measure, for notions like integrability and negligibility, one needs to point out the measure associated. So, π-integrable means "integrable with respect to the measure π".

b) It is clear from 6.15 that

6.17 $$\pi(A \times B) = \mu(A)\nu(B), \qquad A \in \mathcal{E}, B \in \mathcal{F},$$

and for this reason we call π the product of μ and ν and we use the notation $\pi = \mu \times \nu$.

c) If both μ and ν are σ-finite, then Theorem 6.11 applies with $K(x,B) = \nu(B)$ and implies that π is the only measure satisfying 6.17. Otherwise, it is possible that there are measures $\hat{\pi}$ satisfying $\hat{\pi}(A\times B) = \mu(A)\nu(B)$ for all A in $\mathcal{E}$ and B in $\mathcal{F}$ but with $\hat{\pi}f$ differing from πf for some positive f in $\mathcal{E}\otimes\mathcal{F}$.

Proof. a) Let πf be defined by the first integral in 6.15. Taking $K(x,B) = \nu(B)$ in Theorem 6.11 shows that this defines a measure π on the product space. Since $\mu = \sum \mu_i$ and $\nu = \sum \nu_j$ for some finite measures μ_i and ν_j, we have

$$\pi f = \sum_i \sum_j \int_E \mu_i(dx) \int_F \nu_j(dy)\, f(x,y) = \sum_{i,j} (\mu_i \times \nu_j) f$$

by Exercise 4.27 and the monotone convergence theorem. Thus, $\pi = \sum_{i,j} \mu_i \times \nu_j$ and, arranging the pairs (i,j) into a sequence, we see that $\pi = \sum \pi_n$ for some sequence of finite measures π_n.

b) To prove the equality of the integrals in 6.15, we start by observing that the second integral is in fact an integral over $F\times E$: defining $\hat{f}:\ F\times E \mapsto \bar{\mathbb{R}}_+$ by $\hat{f}(y,x) = f(x,y)$, the second integral is

$$\begin{aligned}\hat{\pi}\hat{f} &= \int_F \nu(dy) \int_E \mu(dx)\, \hat{f}(y,x) = \sum_j \sum_i \int_F \nu_j(dy) \int_E \mu_i(dx)\, \hat{f}(y,x) \\ &= \sum_{i,j} (\nu_j \times \mu_i)\, \hat{f}.\end{aligned}$$

Hence, to prove that $\pi f = \hat{\pi}\hat{f}$, it is sufficient to show that $(\mu_i \times \nu_j) f = (\nu_j \times \mu_i)\hat{f}$ for each pair of i and j. Fixing i and j, this amounts to showing that

$$\pi f = (\mu \times \nu) f = (\nu \times \mu) \hat{f} = \hat{\pi}\hat{f}$$

under the assumption that μ and ν are both finite.

c) Assume μ and ν finite. Let $h:\ E\times F \mapsto F\times E$ be the transposition mapping $(x,y) \mapsto (y,x)$. It is obviously measurable relative to $\mathcal{E}\otimes\mathcal{F}$ and $\mathcal{F}\otimes\mathcal{E}$. For sets A in $\mathcal{E}$ and B in $\mathcal{F}$,

$$\pi\circ h^{-1}(B\times A) = \pi(A\times B) = \mu(A)\nu(B) = \hat{\pi}(B\times A),$$

which implies via Proposition 3.7 that $\hat{\pi} = \pi\circ h^{-1}$. Hence, $\hat{\pi}\hat{f} = (\pi\circ h^{-1})\hat{f} = \pi(\hat{f}\circ h) = \pi f$ since $\hat{f}\circ h(x,y) = \hat{f}(y,x) = f(x,y)$.

d) Let f be π-integrable. Then 6.15 holds for f^+ and f^- separately, and $\pi f = \pi f^+ - \pi f^-$ with both terms finite. Hence, 6.15 holds for f. As to the integrability of sections, we observe that the integrability of f implies that $x \mapsto \int_F \nu(dy)\, f(x,y)$ is real-valued for μ-almost every x, which in turn is equivalent to saying that $y \mapsto f(x,y)$ is ν-integrable for μ-almost every x. By symmetry, the finiteness for the second integral implies that $x \mapsto f(x,y)$ is μ-integrable for ν-almost every y.

Finite products

The concepts and results above extend easily to products of finitely many spaces. Let $(E_1, \mathcal{E}_1)$, ..., $(E_n, \mathcal{E}_n)$ be measurable spaces. Their *product* is denoted by any of the following three:

6.18 $$\bigotimes_{i=1}^{n}(E_i, \mathcal{E}_i) = (\bigtimes_{i=1}^{n} E_i, \bigotimes_{i=1}^{n} \mathcal{E}_i) = (E_1 \times \cdots \times E_n, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n),$$

where $E_1 \times \cdots \times E_n$ is the set of all n-tuples $(x_1, \ldots, x_n)$ with x_i in E_i for $i = 1, \ldots, n$, and $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$ is the σ-algebra generated by the *measurable rectangles* $A_1 \times \cdots \times A_n$ with A_i in $\mathcal{E}_i$, $i = 1, \ldots n$.

Let $\mu_1, \ldots, \mu_n$ be Σ-finite measures on $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectively. Then, their *product* $\pi = \mu_1 \times \cdots \times \mu_n$ is the measure defined on the measurable product space by analogy with Theorem 6.14: for positive functions f in $\bigotimes \mathcal{E}_i$,

6.19 $$\pi f = \int_{E_1} \mu_1(dx_1) \int_{E_2} \mu_2(dx_2) \cdots \int_{E_n} \mu_n(dx_n)\, f(x_1, \ldots, x_n).$$

It is usual to denote the resulting measure space

6.20 $$\bigotimes_{i=1}^{n}(E_i, \mathcal{E}_i, \mu_i).$$

Fubini's theorem is generalized to this space and shows that, if f is positive or π-integrable, the integrals on the right side of 6.19 can be performed in any order desired.

More general measures can be defined on the product space 6.18 with the help of kernels. We illustrate the technique for $n = 3$: Let μ_1 be a measure on $(E_1, \mathcal{E}_1)$, let K_2 be a transition kernel from $(E_1, \mathcal{E}_1)$ into $(E_2, \mathcal{E}_2)$, and let K_3 be a transition kernel from $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ into $(E_3, \mathcal{E}_3)$. Consider the formula

6.21 $$\pi f = \int_{E_1} \mu_1(dx_1) \int_{E_2} K_2(x_1, dx_2) \int_{E_3} K_3((x_1, x_2), dx_3)\, f(x_1, x_2, x_3)$$

for positive f in $\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3$. Assuming that K_2 and K_3 are Σ-finite, repeated applications of Theorem 6.11 show that this defines a measure π on $(E_1 \times E_2 \times E_3, \mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3)$.

In situations like this, we shall omit as many parentheses as we can and use a notation analogous to 5.8. For instance, instead of 6.21, we write

6.22 $$\pi(dx_1, dx_2, dx_3) = \mu_1(dx_1)\, K_2(x_1, dx_2)\, K_3(x_1, x_2, dx_3).$$

The notation

6.23 $$\pi = \mu_1 \times K_2 \times K_3$$

is also used for the same thing and is in accord with the notation for product measures.

Infinite products

Let T be an arbitrary set, countable or uncountable. It will play the role of an index set; we think of it as the time set. For each t in T, let $(E_t, \mathcal{E}_t)$ be a measurable space. Let x_t be a point in E_t for each t in T. Then we write $(x_t)_{t\in T}$ for the resulting collection and think of it as a function on T; this is especially appropriate when $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ for all t, because, then, $x = (x_t)_{t\in T}$ can be regarded as the mapping $t \mapsto x_t$ from T into E. The set F of all such functions $x = (x_t)_{t\in T}$ is called the *product space* defined by $\{E_t :\ t \in T\}$; and the notation $\times_{t\in T} E_t$ is used for F.

A *rectangle* in F is a subset of the form

6.24 $$\underset{t\in T}{\times} A_t = \{x \in F :\ x_t \in A_t \text{ for each } t \text{ in } T\}$$

where A_t differs from E_t for only a finite number of t. It is said to be measurable if $A_t \in \mathcal{E}_t$ for every t (for which A_t differs from E_t). The σ-algebra on F generated by the collection of all measurable rectangles is called the product σ-algebra and is denoted by $\bigotimes_{t\in T} \mathcal{E}_t$. The resulting measurable space is denoted variously by

6.25 $$\bigotimes_{t\in T}(E_t, \mathcal{E}_t) = (\underset{t\in T}{\times} E_t, \bigotimes_{t\in T} \mathcal{E}_t).$$

In the special case where $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ for all t, the following notations are also in use for the same:

6.26 $$(E, \mathcal{E})^T = (E^T, \mathcal{E}^T)$$

Although this is the logical point to describe the construction of measures on the product space, we shall delay it until the end of Chapter IV, at which point the steps involved should look intuitive. For the present, we list the following proposition which allows an arbitrary collection of measurable functions to be thought as one measurable function. It is a many-dimensional generalization of the result in Exercise 2.21.

6.27 PROPOSITION. *Let $(\Omega, \mathcal{H})$ be a measurable space. Let $(F, \mathcal{F}) = \otimes_{t\in T}(E_t, \mathcal{E}_t)$. For each t in T, let f_t be a mapping from Ω into E_t. For each ω in Ω, define $f(\omega)$ to be the point $(f_t(\omega))_{t\in T}$ in F. Then, the mapping $f : \Omega \mapsto F$ is measurable relative to $\mathcal{H}$ and $\mathcal{F}$ if and only if f_t is measurable relative to $\mathcal{H}$ and $\mathcal{E}_t$ for every t in T.*

Proof. Suppose that f is measurable relative to $\mathcal{H}$ and $\mathcal{F}$. Then, $\{f \in B\} \in \mathcal{H}$ for every B in $\mathcal{F}$. In particular, taking B to be the rectangle in 6.24 with $A_t = E_t$ for all t except $t = s$ for some fixed s, we see that $\{f \in B\} = \{f_s \in A_s\} \in \mathcal{H}$ for A_s in $\mathcal{E}_s$. Thus, f_s is measurable relative to $\mathcal{H}$ and $\mathcal{E}_s$ for every s fixed.

Suppose that each f_t is measurable relative to $\mathcal{H}$ and $\mathcal{E}_t$. If B is a measurable rectangle in F, then $\{f \in B\}$ is the intersection of finitely many sets of the form $\{f_t \in A_t\}$ with A_t in $\mathcal{E}_t$, and hence, $\{f \in B\} \in \mathcal{H}$. Since measurable rectangles generate the product σ-algebra $\mathcal{F}$, this implies via Proposition 2.3 that f is measurable relative to $\mathcal{H}$ and $\mathcal{F}$. □

Exercises

6.28 *Measurability in the product space.* Suppose that $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, and let $(F, \mathcal{F})$ be arbitrary. Let $f : E \times F \mapsto \bar{\mathbb{R}}$ be such that $y \mapsto f(x, y)$ is $\mathcal{F}$-measurable for each x in E and that $x \mapsto f(x, y)$ is right-continuous (or left-continuous) for each y in F. Show that, then, f is in $\mathcal{E} \otimes \mathcal{F}$.

6.29 *Image measures and kernels.* Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Let $h : E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$. Define

$$K(x, B) = 1_B \circ h(x), \qquad x \in E, B \in \mathcal{F}.$$

Show that K is a transition probability kernel. Show that, in the measure-kernel-function notation of Theorem 6.3,

$$Kf = f \circ h, \quad \mu K = \mu \circ h^{-1}, \quad \mu K f = \mu(f \circ h).$$

6.30 *Transition densities.* Let ν be a σ-finite measure on $(F, \mathcal{F})$, and let k be a positive function in $\mathcal{E} \otimes \mathcal{F}$. Define K by 6.2, that is, in differential notation,

$$K(x, dy) = \nu(dy)\, k(x, y).$$

Show that K is a transition kernel. Then, k is called the transition density function of K with respect to ν.

6.31 *Finite spaces.* Let $E = \{1, \ldots, m\}$, $F = \{1, \ldots, n\}$, $G = \{1, \ldots, p\}$ with their discrete σ-algebras. Functions on such spaces can be regarded as column vectors, measures as row vectors, and kernels as matrices. Show that, with these interpretations, the notations Kf, μK, $\mu K f$, KL used in Theorem 6.3 and Definition 6.5 are in accord with the usual notations used in linear algebra.

6.32 *Finite and bounded kernels.* Let K be a finite transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Define

$$h(x) = \begin{cases} K(x, F) & \text{if } K(x, F) > 0, \\ 1 & \text{if } K(x, F) = 0, \end{cases}$$

and define H by solving

$$K(x, B) = h(x)\, H(x, B).$$

Show that $h \in \mathcal{E}_+$ and that H is a bounded kernel.

6.33 *Proof of Proposition 6.9.* Complete the proof. Hint: Use the preceding exercise to extend the proof from the bounded kernels to finite ones, and finally extend it to Σ-finite kernels.

6.34 *Fubini and Σ-finiteness.* In general, in order for 6.15 to hold, it is necessary that μ and ν be Σ-finite. For instance, let $E = F = [0, 1]$ with their Borel σ-algebras, and let μ be the Lebesgue measure on E, and ν the counting measure on F (that is, $\nu(A)$ is the number of points in A). Then, for $f(x, y) = 1$ if $x = y$ and 0 otherwise, the first integral in 6.15 is equal to 1, but the second is equal to 0.

Complements

6.35 *Product and Borel σ-algebras.* For each t in some index set T, let E_t be a topological space and let $\mathcal{E}_t = \mathcal{B}(E_t)$, the Borel σ-algebra on E_t. Let $(F, \mathcal{F}) = \bigotimes_T (E_t, \mathcal{E}_t)$ be the product measurable space. The product space F can be given the product topology, and let $\mathcal{B}(F)$ be the Borel σ-algebra corresponding to that topology on F.

In general, $\mathcal{B}(F) \supset \mathcal{F}$. If T is countable and if every $\mathcal{E}_t$ has a countable open base, then $\mathcal{F} = \mathcal{B}(F)$. In particular, $\mathbb{R}^n$ and $\mathbb{R}^\infty = \mathbb{R}^\mathbb{N}$ are topological spaces and their Borel σ-algebras coincide with the appropriate product σ-algebras; more precisely

$$(\mathcal{B}_\mathbb{R})^T = \mathcal{B}(\mathbb{R}^T)$$

for $T = \{1, 2, \ldots, n\}$ for every integer $n \geq 1$ and also for $T = \mathbb{N}^* = \{1, 2, \ldots\}$. This equality fails when T is uncountable, $\mathcal{B}(\mathbb{R}^T)$ being the larger then.

6.36 *Standard measurable spaces.* Let $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2), \ldots$ be standard measurable spaces, and let $(F, \mathcal{F})$ be their product. Then, $(F, \mathcal{F})$ is also standard.

Chapter II

Probability Spaces

A probability space is a triplet $(\Omega, \mathcal{H}, \mathbb{P})$ where Ω is a set, $\mathcal{H}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{H})$. Thus, mathematically, a probability space is a special measure space where the measure has total mass one.

But, our attitude and emotional response toward one is entirely different from those toward the other. On a measure space everything is deterministic and certain, on a probability space we face randomness and uncertainty.

A probability space $(\Omega, \mathcal{H}, \mathbb{P})$ is a mathematical model of a random experiment, an experiment whose exact outcome cannot be told in advance. The set Ω stands for the collection of all possible outcomes of the experiment. A subset H is said to *occur* if the outcome of the experiment happens to belong to H. Given our capabilities to measure, detect, and discern, and given the nature of answers we seek, only certain subsets H are distinguished enough to be of concern whether they occur. The σ-algebra $\mathcal{H}$ is the collection of all such subsets whose occurrence are noteworthy and decidable; the elements of $\mathcal{H}$ are called *events*. From this point of view, the conditions for $\mathcal{H}$ to be a σ-algebra are logical consequences of the interpretation of the term "event". Finally, for each event H, the chances that H occurs is modeled to be the number $\mathbb{P}(H)$, called the probability that H occurs.

The actual assignment of probabilities to events is the primary task of the probabilist. It requires much thought and experience, it is rarely explicit, and it determines the quality of the probability space as a model of the experiment involved. Once the probability space is fixed, the main task is to evaluate various integrals of interest by making adroit use of those implicitly defined probabilities. Often, the results are compared against experience, and the probability space is altered for a better fit.

Our aim in this chapter is to introduce the language and notation of probability theory. Implicit in the language are whole sets of attitudes, prejudices, and desires with which we hope to infect the reader.

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_2,

1 Probability Spaces and Random Variables

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. The set Ω is called the *sample space*; its elements are called *outcomes*. The σ-algebra $\mathcal{H}$ may be called the grand *history*; its elements are called *events*. We repeat the properties of the probability measure $\mathbb{P}$; all sets here are events:

1.1 *Norming:* $\mathbb{P}(\emptyset) = 0, \qquad \mathbb{P}(\Omega) = 1.$

Monotonicity: $H \subset K \quad \Rightarrow \quad \mathbb{P}(H) \leq \mathbb{P}(K).$

Finite additivity: $H \cap K = \emptyset \quad \Rightarrow \quad \mathbb{P}(H \cup K) = \mathbb{P}(H) + \mathbb{P}(K).$

Countable additivity: (H_n) disjointed $\quad \Rightarrow \quad \mathbb{P}(\bigcup_n H_n) = \sum_n \mathbb{P}(H_n).$

Sequential continuity: $H_n \nearrow H \quad \Rightarrow \quad \mathbb{P}(H_n) \nearrow \mathbb{P}(H),$
$H_n \searrow H \quad \Rightarrow \quad \mathbb{P}(H_n) \searrow \mathbb{P}(H).$

Boole's inequality: $\mathbb{P}(\bigcup_n H_n) \quad \leq \quad \sum_n \mathbb{P}(H_n).$

All of these are as before for arbitrary measures, except for the sequential continuity under decreasing limits, which is made possible by the finiteness of $\mathbb{P}$: If $H_1 \supset H_2 \supset \ldots$ and $\lim H_n = \bigcap H_n = H$, then the complements H_n^c increase to H^c, which implies that $\mathbb{P}(H_n^c) \nearrow \mathbb{P}(H^c)$ by the sequential continuity of measures under increasing limits, and we have $\mathbb{P}(H) = 1 - \mathbb{P}(H^c)$, and similarly for each H_n, by the finite additivity and norming of $\mathbb{P}$.

Negligibility, completeness

The concepts are the same as for arbitrary measures: A subset N of Ω is said to be *negligible* if there exists an event H such that $N \subset H$ and $\mathbb{P}(H) = 0$. The probability space is said to be *complete* if every negligible set is an event.

Improbable events do not bother the probabilist. Negligible sets should not either, but if a negligible set does not belong to $\mathcal{H}$ then we are not able to talk of its probability, which thing is bothersome. So, it is generally nicer to have $(\Omega, \mathcal{H}, \mathbb{P})$ complete. If it is not, it can be completed using Proposition I.3.10.

Almost surely, almost everywhere

An event is said to be *almost sure* if its probability is one. If a proposition holds for every outcome ω in an almost sure event, then we say that the proposition holds *almost surely* or *almost everywhere* or *for almost every* ω or *with probability one*. Obviously, the concept is equivalent to having the proposition fail only over a negligible set.

Random variables

Let $(E, \mathcal{E})$ be a measurable space. A mapping $X: \Omega \mapsto E$ is called a *random variable* taking values in $(E, \mathcal{E})$ provided that it be measurable relative to $\mathcal{H}$ and $\mathcal{E}$, that is, if

1.2 $$X^{-1}A = \{X \in A\} = \{\omega \in \Omega: X(\omega) \in A\}$$

is an event for every A in $\mathcal{E}$. Of course, it is sufficient to check the condition for A in a collection that generates $\mathcal{E}$. It is customary to denote random variables by capital letters.

If the σ-algebra $\mathcal{E}$ is understood from context, then we merely say that X takes values in E or that X is *E-valued*. This is especially the case if E is $\mathbb{R}$ or $\mathbb{R}^d$ or some Borel subset of some such space and $\mathcal{E}$ is the Borel σ-algebra on E.

The simplest random variables are indicators of events; we use the usual notation 1_H for the indicator of H. A random variable is *simple* if it takes only finitely many values, all in $\mathbb{R}$. It is said to be *discrete* if it is elementary, that is, if it takes only countably many values.

Distribution of a random variable

Let X be a random variable taking values in some measurable space $(E, \mathcal{E})$. Let μ be the image of $\mathbb{P}$ under X (see section I.5 for image measures), that is,

1.3 $$\mu(A) = \mathbb{P}(X^{-1}A) = \mathbb{P}\{X \in A\}, \qquad A \in \mathcal{E},$$

where the last member is read as "the probability that X is in A". Then, μ is a probability measure on $(E, \mathcal{E})$; it is called the *distribution* of X.

In view of Proposition I.3.7, to specify the distribution μ, it is sufficient to specify $\mu(A)$ for all A belonging to a p-system that generates $\mathcal{E}$. In particular, if $E = \bar{\mathbb{R}}$ and $\mathcal{E} = \mathcal{B}_E$, the intervals $[-\infty, x]$ with x in $\mathbb{R}$ form a convenient p-system; consequently, in this case, it is enough to specify

1.4 $$c(x) = \mu[-\infty, x] = \mathbb{P}\{X \le x\}, \qquad x \in \mathbb{R}.$$

The resulting function $c: \mathbb{R} \mapsto [0, 1]$ is called the *distribution function* of X. Distribution functions are used extensively in elementary probability theory in order to avoid measures. We shall have little use for them. A review of some salient facts are put as exercises for the sake of completeness.

Functions of random variables

Let X be a random variable taking values in $(E, \mathcal{E})$. Let $(F, \mathcal{F})$ be another measurable space, and let $f: E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$. Then, the composition $Y = f \circ X$ of X and f, namely,

1.5 $$Y(\omega) = f \circ X(\omega) = f(X(\omega)), \qquad \omega \in \Omega,$$

is a random variable taking values in $(F, \mathcal{F})$; this follows from Proposition I.2.5 that measurable functions of measurable functions are measurable. If μ is the distribution of X, then the distribution ν of Y is $\nu = \mu \circ f^{-1}$:

1.6 $$\nu(B) = \mathbb{P}\{Y \in B\} = \mathbb{P}\{X \in f^{-1}B\} = \mu(f^{-1}B), \qquad B \in \mathcal{F}.$$

Joint distributions

Let X and Y be random variables taking values in measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively. Then, the pair $Z = (X, Y) : \omega \mapsto Z(\omega) = (X(\omega), Y(\omega))$ is measurable relative to $\mathcal{H}$ and the product σ-algebra $\mathcal{E} \otimes \mathcal{F}$, that is, Z is a random variable taking values in the product space $(E \times F,\ \mathcal{E} \otimes \mathcal{F})$.

The distribution of Z is a probability measure π on the product space and is also called the *joint distribution* of X and Y. Since $\mathcal{E} \otimes \mathcal{F}$ is generated by the p-system of measurable rectangles, in order to specify π it is sufficient to specify

1.7 $$\pi(A \times B) = \mathbb{P}\{X \in A, Y \in B\}, \qquad A \in \mathcal{E}, B \in \mathcal{F},$$

the right side being the probability that X is in A and Y is in B, that is, the probability of $\{X \in A\} \cap \{Y \in B\}$. In the opposite direction, given the joint distribution π, for A in $\mathcal{E}$ and B in $\mathcal{F}$, we have

1.8 $$\mu(A) = \mathbb{P}\{X \in A\} = \pi(A \times F), \qquad \nu(B) = \mathbb{P}\{Y \in B\} = \pi(E \times B).$$

In this context, the probability measures μ and ν are called the *marginal distributions* of X and Y respectively. These terms are used, with obvious generalizations, for any finite number of random variables.

Independence

Let X and Y be random variables taking values in $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively, and let μ and ν be their respective (marginal) distributions. Then, X and Y are said to be *independent* if their joint distribution is the product measure formed by their marginals, that is, if the distribution of the pair (X, Y) is the product measure $\mu \times \nu$, or in still other words,

1.9 $$\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}, \qquad A \in \mathcal{E}, B \in \mathcal{F}.$$

In probability theory, independence is used often as a primitive concept to be decided by considerations based on the underlying experiment and the way X and Y are defined. And, once it is decided upon, independence of X and Y becomes a convenient tool for specifying the joint distribution via its marginals. We shall return to these matters in Chapter IV for a rigorous treatment. For the present we mention an extension or two.

A finite collection $\{X_1, \ldots, X_n\}$ of random variables is said to be an *independency*, or the variables $X_1, \ldots, X_n$ are said to be *independent*, if the distribution of the random vector $(X_1, \ldots, X_n)$ has the product form $\mu_1 \times \cdots \times \mu_n$ where $\mu_1, \ldots, \mu_n$ are probability measures. Then, necessarily, μ_i is the distribution of X_i for each i. An arbitrary collection (countable or uncountable) of random variables is said to be an *independency* if every finite sub-collection of it is an independency.

Stochastic processes and probability laws

Let $(E, \mathcal{E})$ be a measurable space. Let T be an arbitrary set, countable or uncountable. For each t in T, let X_t be a random variable taking values in $(E, \mathcal{E})$. Then, the collection $\{X_t : \ t \in T\}$ is called a *stochastic process* with *state space* $(E, \mathcal{E})$ and *parameter set* T.

For each ω in Ω, let $X(\omega)$ denote the function $t \mapsto X_t(\omega)$ from T into E; then, $X(\omega)$ is an element of E^T. By Proposition I.6.27, the mapping $X : \ \omega \mapsto X(\omega)$ from Ω into E^T is measurable relative to $\mathcal{H}$ and $\mathcal{E}^T$. In other words, we may regard the stochastic process $\{X_t : \ t \in T\}$ as a random variable X that takes values in the product space $(F, \mathcal{F}) = (E^T, \mathcal{E}^T)$.

The distribution of the random variable X, that is, the probability measure $\mathbb{P} \circ X^{-1}$ on $(F, \mathcal{F})$, is called the *probability law* of the stochastic process $\{X_t : \ t \in T\}$.

Recall that the product σ-algebra $\mathcal{F}$ is generated by the finite-dimensional rectangles and, therefore, a probability measure on $(F, \mathcal{F})$ is determined by the values it assigns to those rectangles. It follows that the probability law of X is determined by the values

1.10 $$\mathbb{P}\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}$$

with n ranging over $\mathbb{N}^*$, and $t_1, \ldots, t_n$ over T, and $A_1, \ldots, A_n$ over $\mathcal{E}$. Much of the theory of stochastic processes has to do with computing integrals concerning X from the given data regarding 1.10.

Examples of distributions

The aim here is to introduce a few distributions that are encountered often in probabilistic work. Other examples will appear in the exercises below and in the section next.

1.11 *Poisson distribution.* Let X be a random variable taking values in $\mathbb{N} = \{0, 1, \ldots\}$; it is to be understood that the relevant σ-algebra on $\mathbb{N}$ is the discrete σ-algebra of all subsets. Then, X is said to have the Poisson distribution with mean c if

$$\mathbb{P}\{X = n\} = \frac{e^{-c} \, c^n}{n!}, \qquad n \in \mathbb{N}.$$

Here, c is a strictly positive real number. The corresponding distribution is the probability measure μ on $\mathbb{N}$ defined by

$$\mu(A) = \sum_{n\in A} \frac{e^{-c}c^n}{n!}\,, \qquad A \subset \mathbb{N}.$$

1.12 *Exponential distributions.* Let X be a random variable with values in $\mathbb{R}_+$; the relevant σ-algebra on $\mathbb{R}_+$ is $\mathcal{B}(\mathbb{R}_+)$. Then, X is said to have the exponential distribution with scale parameter c if its distribution μ has the form

$$\mu(dx) = dx\; ce^{-cx}\,, \quad x \in \mathbb{R}_+,$$

where dx is short for $\text{Leb}(dx)$. Here, $c > 0$ is a constant, and we used the form I.5.8 to display μ. In other words, μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$ and its density function is $p(x) = ce^{-cx}$, $x \in \mathbb{R}_+$. When $c = 1$, this distribution is called the standard exponential.

1.13 *Gamma distributions.* Let X be a random variable with values in $\mathbb{R}_+$. It is said to have the gamma distribution with shape index a and scale parameter c if its distribution μ has the form

$$\mu(dx) = dx\; \frac{c^a x^{a-1} e^{-cx}}{\Gamma(a)}\,, \qquad x \in \mathbb{R}_+.$$

Here, $a > 0$ and $c > 0$ are constants and $\Gamma(a)$ is the so-called gamma function. The last is defined so that μ is a probability measure, that is,

$$\Gamma(a) = \int_0^\infty dx\; x^{a-1}e^{-x}.$$

Incidentally, the density function for μ takes the value $+\infty$ at $x = 0$ if $a < 1$, but this is immaterial since $\text{Leb}\{0\} = 0$; or, in probabilistic terms, $X \in \mathbb{R}_+^* = (0, \infty)$ almost surely, and it is sufficient to define the density on $\mathbb{R}_+^*$. In general, $\Gamma(a) = (a-1)\Gamma(a-1)$ for $a > 1$. This allows one, together with $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(1) = 1$, to give an explicit expression for $\Gamma(a)$ when $a > 0$ is an integer or half-integer. In particular, when $a = 1$, the gamma distribution becomes the exponential; and when $c = \frac{1}{2}$ and $a = \frac{n}{2}$ for some integer $n \geq 1$, it is also called the Chi-square distribution with n degrees of freedom. Finally, when $c = 1$, we call the distribution standard gamma distribution with shape index a.

1.14 *Gaussian distributions.* Let X be a real-valued random variable. It is said to have the Gaussian (or normal) distribution with mean a and variance b if its distribution μ has the form

$$\mu(dx) = dx\; \frac{1}{\sqrt{2\pi b}}\, e^{-(x-a)^2/2b}, \qquad x \in \mathbb{R}.$$

Here, $a \in \mathbb{R}$ and $b > 0$, both constant. If $a = 0$ and $b = 1$, then μ is called the standard Gaussian distribution.

1.15 *Independent gamma variables.* Let γ_a denote the standard gamma distribution with shape index a; this is the probability measure μ of Example 1.13 above but with $c = 1$. Let X have the distribution γ_a, and Y the distribution γ_b; here $a > 0$ and $b > 0$. Suppose that X and Y are independent. Then, the joint distribution of X and Y is the product measure $\gamma_a \times \gamma_b$, that is, the distribution of the pair (X, Y) is the probability measure π on $\mathbb{R}_+ \times \mathbb{R}_+$ given by

$$\pi(dx, dy) = \gamma_a(dx)\ \gamma_b(dy) = dx\ dy\ \frac{e^{-x}\ x^{a-1}}{\Gamma(a)} \cdot \frac{e^{-y}\ y^{b-1}}{\Gamma(b)}.$$

1.16 *Gaussian with exponential variance.* Let X and Y be random variables taking values in $\mathbb{R}_+$ and $\mathbb{R}$ respectively. Suppose that their joint distribution π is given by

$$\pi(dx, dy) = dx\ dy\ ce^{-cx}\ \frac{1}{\sqrt{2\pi x}}\ e^{-y^2/2x}\ , \qquad x \in \mathbb{R}_+,\ y \in \mathbb{R}.$$

Note that π has the form $\pi(dx, dy) = \mu(dx)\ K(x, dy)$, where μ is the exponential distribution with scale parameter c, and for each x, the distribution $B \mapsto K(x, B)$ is Gaussian with mean 0 and variance x. Indeed, K is a transition kernel from $\mathbb{R}_+$ into $\mathbb{R}$, and π is an instance of the measure appearing in Theorem I.6.11. It is clear that the marginal distribution of X is the exponential distribution μ. The marginal distribution ν of Y has the form $\nu = \mu K$ introduced in Theorem I.6.3:

$$\nu(B) = \pi(\mathbb{R}_+ \times B) = \int_{\mathbb{R}_+} \mu(dx)\ K(x, B)\ , \qquad B \in \mathcal{B}_{\mathbb{R}}.$$

It is seen easily that ν is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, that is, ν has the form $\nu(dy) = dy \cdot n(y)$, and the density function is

$$n(y) = \int_0^\infty dx\ ce^{-cx}\ \frac{e^{-y^2/2x}}{\sqrt{2\pi x}} = \tfrac{1}{2}\ b\ e^{-b|y|}\ , \qquad y \in \mathbb{R},$$

with $b = \sqrt{2c}$. Incidentally, this distribution ν is called the *two-sided exponential* distribution with parameter b. Finally, we note that π is not the product $\mu \times \nu$, that is, X and Y are dependent variables.

Exercises and complements

1.17 *Distribution functions.* Let X be a random variable taking values in $\bar{\mathbb{R}} = [-\infty, +\infty]$. Let μ be its distribution, and c its distribution function, defined by 1.4. Then, c is a function from $\mathbb{R}$ into $[0, 1]$. It is increasing and right-continuous as indicated in Exercise I.5.14.

a) Since c is increasing, the left-hand limit

$$c(x-) = \lim_{y \uparrow x} c(y)$$

exists for every x in $\mathbb{R}$. Similarly, the limits

$$c(-\infty) = \lim_{x \downarrow -\infty} c(x) \qquad c(+\infty) = \lim_{x \uparrow \infty} c(x)$$

exist. Show that

$$c(x-) = \mathbb{P}\{X < x\}, \qquad c(x) - c(x-) = \mathbb{P}\{X = x\}$$

$$c(-\infty) = \mathbb{P}\{X = -\infty\}, \qquad c(+\infty) = \mathbb{P}\{X < \infty\} = 1 - \mathbb{P}\{X = \infty\}.$$

b) Let D be the set of all atoms of the distribution μ. Then, D consists of all x in $\mathbb{R}$ for which $c(x) - c(x-) > 0$, plus the point $-\infty$ if $c(-\infty) > 0$, plus the point $+\infty$ if $c(\infty) < 1$. Of course, D is countable. Define $D_x = D \cap (-\infty, x]$ and

$$a(x) = c(-\infty) + \sum_{y \in D_x} [c(y) - c(y-)] \, , \qquad b(x) = c(x) - a(x)$$

for x in $\mathbb{R}$. Then, a is an increasing right-continuous function that increases by jumps only, and b is increasing continuous. Show that a is the distribution function of the measure

$$\mu_a(B) = \mu(B \cap D) \, , \qquad B \in \mathcal{B}(\bar{\mathbb{R}}),$$

and b is the distribution function of the measure $\mu_b = \mu - \mu_a$. Note that μ_a is purely atomic and μ_b is diffuse. The random variable X is almost surely discrete if and only if $\mu = \mu_a$, that is, $a = c$.

1.18 *Quantile functions.* Let X be real-valued, let c be its distribution function. Note that, then, $c(-\infty) = 0$ and $c(+\infty) = 1$. Suppose that c is continuous and strictly increasing, and let q be the functional inverse of c, that is, $q(u) = x$ if and only if $c(x) = u$ for u in $(0, 1)$. The function $q : \ (0,1) \mapsto \mathbb{R}$ is called the quantile function of X since

$$\mathbb{P}\{X \le q(u)\} = u \, , \qquad u \in (0,1).$$

Let U be a random variable having the *uniform distribution* on $(0,1)$, that is, the distribution of U is the Lebesgue measure on $(0,1)$. Show that, then, the random variable $Y = q \circ U$ has the same distribution as X. In general, $Y \ne X$.

1.19 *Continuation.* This is to re-do the preceding exercise assuming that $c : \mathbb{R} \mapsto [0,1]$ is only increasing and right-continuous. Let $q : \ (0,1) \mapsto \bar{\mathbb{R}}$ be the right-continuous functional inverse of c, that is,

$$q(u) = \inf\{x \in \mathbb{R} : \quad c(x) > u\}$$

with the usual conventions that $\inf \mathbb{R} = -\infty$, $\inf \emptyset = +\infty$. We call q the quantile function corresponding to c by analogy with the preceding exercise. Recall from Exercise I.5.13 that q is increasing and right-continuous, and that c is related to q by the same formula with which q is related to c. Note that q is real-valued if and only if $c(-\infty) = 0$ and $c(+\infty) = 1$. See also Figure 1. Show that $c(x-) \le u$ if and only if $q(u) \ge x$, and, by symmetry, $q(u-) \le x$ if and only if $c(x) \ge u$.

1.20 *Construction of probability measures on* $\bar{\mathbb{R}}$. Let c be a cumulative distribution function, that is, $c : \mathbb{R} \mapsto [0,1]$ is increasing and right-continuous. Let $q : (0,1) \mapsto \bar{\mathbb{R}}$ be the corresponding quantile function. Let λ denote the Lebesgue measure on $(0,1)$ and put $\mu = \lambda \circ q^{-1}$. Show that μ is a probability measure on $\bar{\mathbb{R}}$. Show that μ is the distribution on $\bar{\mathbb{R}}$ corresponding to the distribution function c. Thus, to every distribution function c on $\mathbb{R}$ there corresponds a unique probability measure μ on $\bar{\mathbb{R}}$ and vice-versa.

1.21 *Construction of random variables.* Let μ be a probability measure on $\bar{\mathbb{R}}$. Then, there exists a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and a random variable $X : \Omega \mapsto \bar{\mathbb{R}}$ such that μ is the distribution of X: Take $\Omega = (0,1)$, $\mathcal{H} = \mathcal{B}_{(0,1)}$, $\mathbb{P} = \text{Leb}$, and define $X(\omega) = q(\omega)$ for ω in Ω, where q is the quantile function corresponding to the measure μ (via the cumulative distribution function). See Exercise I.5.15 for the extension of this construction to abstract spaces. This setup is the theoretical basis of Monte-Carlo studies.

1.22 *Supplement on quantiles.* Literature contains definitions similar to that in 1.19 for q, but with slight differences, one of the popular ones being

$$p(u) = \inf\{x \in \mathbb{R} : \quad c(x) \ge u\}\ , \qquad u \in (0,1).$$

Some people prefer supremums, but there is nothing different, since $q(u) = \sup\{x : \ c(x) \le u\}$ and $p(u) = \sup\{x : \ c(x) < u\}$. In fact, there is close relationship between p and q: we have $p(u) = q(u-) = \lim_{v \nearrow u} q(v)$. The function q is right-continuous, whereas p is left-continuous. We prefer q over p, because q and c are functional inverses of each other. Incidentally, in the constructions of 1.20 and 1.21 above, the minor difference between p and q proves unimportant: Since q is increasing and right-continuous, $p(u) = q(u-)$ differs from $q(u)$ for at most countably many u; therefore, $\text{Leb}\{u : p(u) \ne q(u)\} = 0$ and, hence, $\lambda \circ q^{-1} = \lambda \circ p^{-1}$ with $\lambda = \text{Leb}$ on $(0,1)$.

2 Expectations

Throughout this section $(\Omega, \mathcal{H}, \mathbb{P})$ is a probability space and all random variables are defined on Ω and take values in $\bar{\mathbb{R}}$, unless stated otherwise.

Let X be a random variable. Since it is $\mathcal{H}$-measurable, its integral with respect to the measure $\mathbb{P}$ makes sense to talk about. That integral is called the *expected value* of X and is denoted by any of the following

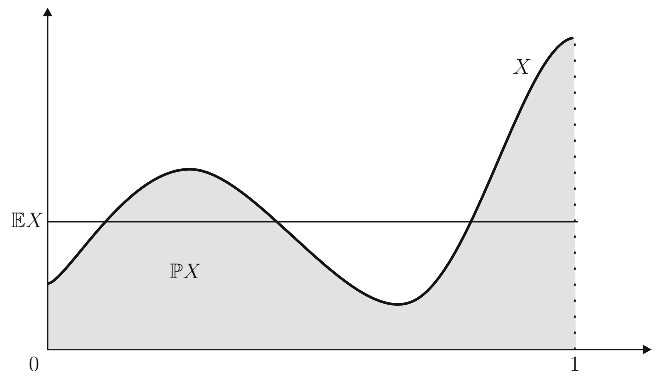


Figure 2: The integral $\mathbb{P}X$ is the area under X, the expected value $\mathbb{E}X$ is the constant "closest" to X.

2.1 $$\mathbb{E}X = \int_\Omega \mathbb{P}(d\omega)\ X(\omega) = \int_\Omega X\ d\mathbb{P} = \mathbb{P}X.$$

The expected value $\mathbb{E}X$ exists if and only if the integral does, that is, if and only if we do not have $\mathbb{E}X^+ = \mathbb{E}X^- = +\infty$. Of course, $\mathbb{E}X$ exists whenever $X \geq 0$, and $\mathbb{E}X$ exists and is finite if X is bounded.

We shall treat $\mathbb{E}$ as an operator, the expectation operator corresponding to $\mathbb{P}$, and call $\mathbb{E}X$ the expectation of X from time to time. The change in notation serves to highlight the important change in our interpretation of $\mathbb{E}X$: The integral $\mathbb{P}X$ is the "area under the function" X in a generalized sense. The expectation $\mathbb{E}X$ is the "weighted average of the values" of X, the weight distribution being specified by $\mathbb{P}$, the total weight being $\mathbb{P}(\Omega) = 1$. See Figure 2 above for the distinction.

Except for this slight change in notation, all the conventions and notations of integration are carried over to expectations. In particular, X is said to be *integrable* if $\mathbb{E}X$ exists and is finite. The integral of X over an event H is $\mathbb{E}X1_H$. As before with integrals, we shall state most results for positive random variables, because expectations exist always for such, and because the extensions to arbitrary random variables are generally obvious.

Properties of expectation

The following is a rapid summary of the main results on integrals stated in probabilistic terms. Here, X, Y, etc. are random variables taking values in $\bar{\mathbb{R}}$, and a, b, etc. are positive constants.

2.2 *Positivity:* $X \geq 0 \Rightarrow \mathbb{E}X \geq 0$.

Monotonicity: $X \geq Y \geq 0 \Rightarrow \mathbb{E}X \geq \mathbb{E}Y$.

Linearity: $X, Y \geq 0 \Rightarrow \mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$.

Insensitivity: $X, Y \geq 0,\ X = Y$ almost surely $\Rightarrow\ \mathbb{E}X = \mathbb{E}Y$.

Monotone convergence: $X_n \geq 0,\ X_n \nearrow X\ \Rightarrow\ \mathbb{E}X_n \nearrow \mathbb{E}X$
$X_n \geq 0,\ \Rightarrow\ \mathbb{E}\sum X_n = \sum \mathbb{E}X_n$.

Fatou's Lemma: $X_n \geq 0\ \Rightarrow\ \mathbb{E}\liminf X_n\ \leq\ \liminf \mathbb{E}X_n$.

Dominated convergence: $|X_n| \leq Y,\ Y$ integrable, $\lim X_n$ exists
$\Rightarrow\ \mathbb{E}\lim X_n = \lim \mathbb{E}X_n$.

Bounded convergence: $|X_n| \leq b,\ b < \infty,\ \lim X_n$ exists
$\Rightarrow\ \mathbb{E}\lim X_n = \lim \mathbb{E}X_n$.

2.3 REMARKS. a) Positivity can be added to: for $X \geq 0$, we have $\mathbb{E}X = 0$ if and only if $X = 0$ almost surely.

b) Monotonicity can be extended: if $X \geq Y$, then $\mathbb{E}\,X \geq \mathbb{E}\,Y$ provided that both $\mathbb{E}\,X$ and $\mathbb{E}\,Y$ exist (infinite values are allowed); if $X \geq Y$ and either $\mathbb{E}X$ or $\mathbb{E}Y$ is finite, then both $\mathbb{E}X$ and $\mathbb{E}Y$ exist and $\mathbb{E}X \geq \mathbb{E}Y$.

c) Insensitivity can be extended likewise: if $X = Y$ almost surely and either $\mathbb{E}X$ or $\mathbb{E}Y$ exists, then so is the other and $\mathbb{E}X = \mathbb{E}Y$.

d) The preceding two remarks have a useful partial converse: If $\mathbb{E}\ X1_H \geq \mathbb{E}\ Y1_H$ for every event H, then $X \geq Y$ almost surely. To show this, we use the remark above on monotonicity and the assumed inequality with $H = \{X < q < r < Y\}$, where q and r are rational numbers with $q < r$. This yields

$$q\mathbb{P}(H) = \mathbb{E}\ q1_H \geq \mathbb{E}\ X1_H \geq \mathbb{E}\ Y1_H \geq \mathbb{E}\ r1_H = r\mathbb{P}(H),$$

which is possible with $q < r$ only if $\mathbb{P}(H) = 0$. Hence, the event $\{Y > X\}$ has probability zero, via Boole's inequality, since it is the union of events like H over all rationals q and r with $q < r$.

e) Convergence theorems have various generalizations along the lines indicated for integrals. For example, an easy consequence of the monotone convergence theorem is that if $X_n \leq Y$ for all n for some integrable Y, and if $X_n \searrow X$, then $\mathbb{E}X_n \searrow \mathbb{E}X$.

f) Convergence theorems have almost sure versions similar to almost everywhere versions with integrals.

g) If a mapping $X:\ \Omega \mapsto \bar{\mathbb{R}}$ is equal to a random variable Y almost surely, and even if $X(\omega)$ is specified only for almost every ω, the expected value of X is defined to be $\mathbb{E}Y$.

Expectations and integrals

The following relates expectations, which are integrals with respect to $\mathbb{P}$, to integrals with respect to distributions. This is the work horse of computations. Recall that $\mathcal{E}_+$ is the collection of all positive $\mathcal{E}$-measurable functions (from E into $\bar{\mathbb{R}}_+$).

2.4 THEOREM. *Let X be a random variable taking values in some measurable space $(E, \mathcal{E})$. If μ is the distribution of X, then*

2.5 $$\mathbb{E}\ f \circ X = \mu f$$

for every f in $\mathcal{E}_+$. Conversely, if 2.5 holds for some measure μ and all f in $\mathcal{E}_+$, then μ is the distribution of X.

Proof. The first statement is a re-phrasing of Theorem I.5.2 on integration with respect to image measures: if $\mu = \mathbb{P} \circ X^{-1}$, then $\mu f = \mathbb{P}(f \circ X) = \mathbb{E}\ f \circ X$ at least for f in $\mathcal{E}_+$. Conversely, if 2.5 holds for all f in $\mathcal{E}_+$, taking $f = 1_A$ in particular, we see that

$$\mu(A) \ = \ \mu\, 1_A \ = \ \mathbb{E}\ 1_A \circ X \ = \ \mathbb{P}\{X \in A\},$$

that is, μ is the distribution of X. □

In the preceding theorem, the restriction to positive f is for reasons of convenience. For f in $\mathcal{E}$, the formula 2.5 holds for f^+ and f^- respectively, and hence for f, provided that either the expectation $\mathbb{E}\ f \circ X$ or the integral μf exists (then so does the other). The converse statement is useful for figuring out the distribution of X in cases where X is a known function of other random variables whose joint distribution is known. In such cases, it encompasses the formula 1.6 and is more intuitive; we shall see several illustrations of its use below.

Obviously, for a measure μ to be the distribution of X it is sufficient to have 2.5 hold for all f having the form $f = 1_A$ with A in $\mathcal{E}$, or with A in some p-system generating $\mathcal{E}$. When E is a metrizable topological space and $\mathcal{E} = \mathcal{B}(E)$, it is also sufficient to have 2.5 hold for all f that are bounded, positive, and continuous; see Exercise 2.36 in this connection.

Means, variances, Laplace and Fourier transforms

Certain expected values have special names. Let X be a random variable taking values in $\bar{\mathbb{R}}$ and having the distribution μ. The expected value of the nth power of X, namely $\mathbb{E}X^n$, is called the nth *moment* of X. In particular, $\mathbb{E}X$ is also called the *mean* of X. Assuming that the mean is finite (that is, X is integrable), say $\mathbb{E}X = a$, the nth moment of $X - a$ is called the nth centered moment of X. In particular, $\mathbb{E}(X-a)^2$ is called the *variance* of X, and we shall denote it by $\text{Var}X$; note that

2.6 $$\text{Var}\ X = \mathbb{E}\ (X-a)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2,$$

assuming of course that $a = \mathbb{E}X$ is finite.

Assuming that X is positive, for r in $\mathbb{R}_+$, the random variable e^{-rX} takes values in the interval $[0,1]$, and its expectation

2.7 $$\hat{\mu}_r = \mathbb{E}\ e^{-rX} = \int_{\mathbb{R}_+} \mu(dx)\ e^{-rx}$$

is a number in $[0, 1]$. The resulting function $r \mapsto \hat{\mu}_r$ from $\mathbb{R}_+$ into $[0, 1]$ is called the *Laplace transform* of the distribution μ, and by an abuse of language, also the Laplace transform of X.

It can be shown that the Laplace transform determines the distribution: if μ and ν are distributions on $\bar{\mathbb{R}}_+$, and $\hat{\mu}_r = \hat{\nu}_r$ for all r in $\mathbb{R}_+$, then $\mu = \nu$: see Exercise 2.36 below.

Suppose that X is real-valued, that is, X takes values in $\mathbb{R}$. For r in $\mathbb{R}$, $e^{irX} = \cos rX + i \sin rX$ is a complex-valued random variable (here $i = \sqrt{-1}$), and the notion of expected value extends to it naturally:

2.8 $$\hat{\mu}_r = \mathbb{E}\ e^{irX} = \int_{\mathbb{R}} \mu(dx)\ e^{irx} = \int_{\mathbb{R}} \mu(dx)\ \cos rx + i \int_{\mathbb{R}} \mu(dx)\ \sin rx.$$

The resulting complex-valued function $r \mapsto \hat{\mu}_r$ from $\mathbb{R}$ into the complex plane is called the *Fourier transform* of the distribution μ, or the *characteristic function* of the random variable X. As with Laplace transforms, the Fourier transform determines the distribution.

Finally, if X takes values in $\bar{\mathbb{N}} = \{0, 1, \ldots, +\infty\}$, then

2.9 $$\mathbb{E}\ z^X = \sum_{n=0}^{\infty} z^n\ \mathbb{P}\{X = n\}\ , \qquad z \in [0, 1],$$

defines a function from $[0, 1]$ into $[0, 1]$ which is called the *generating function* of X. It determines the distribution of X: in a power series expansion of it, the coefficient of z^n is $\mathbb{P}\{X = n\}$ for each n in $\mathbb{N}$.

Examples

2.10 *Gamma distribution.* Fix $a > 0$ and $c > 0$, and let $\gamma_{a,c}$ be the gamma distribution with shape index a and scale parameter c; see Example 1.13. Let X have $\gamma_{a,c}$ as its distribution. Then, X has finite moments of all orders. Indeed, for every p in $\mathbb{R}_+$,

$$\begin{aligned}\mathbb{E}\ X^p &= \int_0^\infty \gamma_{a,c}(dx)\ x^p = \int_0^\infty dx\ \frac{c^a\ x^{a-1}\ e^{-cx}}{\Gamma(a)}\ x^p \\ &= \frac{\Gamma(a+p)}{c^p\ \Gamma(a)} \int_0^\infty dx\ \frac{c^{a+p}\ x^{a+p-1}\ e^{-cx}}{\Gamma(a+p)} = \frac{\Gamma(a+p)}{\Gamma(a)}\ c^{-p},\end{aligned}$$

since the last integral is $\gamma_{a+p,c}(\mathbb{R}_+) = 1$. Finally, to explain the term "scale parameter" for c, we show that cX has the standard gamma distribution with shape index a (to understand the term "shape index" draw the density function of γ_a for $a < 1, a = 1, a > 1$). To this end, we use Theorem 2.4. Let f be a positive Borel function on $\mathbb{R}_+$. Then,

$$\mathbb{E}f(cX) = \int_0^\infty dx\ \frac{c^a\ x^{a-1}\ e^{-cx}}{\Gamma(a)}\ f(cx) = \int_0^\infty dy\ \frac{y^{a-1}\ e^{-y}}{\Gamma(a)}\ f(y),$$

which means that cX has the distribution γ_a, the standard gamma with shape index a.

2.11 *Gamma and gamma and beta.* Let X and Y be as in Example 1.15, that is, X and Y are independent, X has the standard gamma distribution γ_a with shape index a, and Y has the standard gamma distribution γ_b with shape index b. We now show that

a) $X + Y$ has the standard gamma distribution γ_{a+b},

b) $X/(X + Y)$ has the distribution

$$\beta_{a,b}(du) = du\, \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, u^{a-1}\, (1-u)^{b-1}\,, \quad 0 < u < 1,$$

which is called the beta distribution with index pair (a, b), and

c) $X + Y$ and $X/(X + Y)$ are independent, that is, their joint distribution π is the product measure $\gamma_{a+b} \times \beta_{a,b}$.

We show all this by using the method of Theorem 2.4. Let f be a positive Borel function on $\mathbb{R}_+ \times [0, 1]$ and consider the integral πf:

$$\begin{aligned}
\pi f &= \mathbb{E}\, f(X+Y, \frac{X}{X+Y}) \\
&= \int_0^\infty dx\, \frac{x^{a-1}\, e^{-x}}{\Gamma(a)} \int_0^\infty dy\, \frac{y^{b-1}\, e^{-y}}{\Gamma(b)}\, f(x+y, \frac{x}{x+y}) \\
&= \int_0^\infty dz \int_0^1 du\, \frac{z^{a+b-1}\, e^{-z}}{\Gamma(a)\, \Gamma(b)}\, u^{a-1}\, (1-u)^{b-1}\, f(z, u),
\end{aligned}$$

where the last line is obtained by replacing x with uz and y with $(1-u)z$, and noting that the Jacobian of the transformation is equal to z. There remains to note that the last expression is equal to $(\gamma_{a+b} \times \beta_{a,b})f$, which proves all three claims together.

2.12 *Laplace transforms and distributions and Pareto.* Let X be a random variable taking values in $\mathbb{R}_+$. Then, the Laplace transform $r \mapsto \mathbb{E}\, e^{-rX}$ is a decreasing continuous function on $\mathbb{R}_+$ with value 1 at $r = 0$. Hence, there is a positive random variable R such that

$$\mathbb{P}\{R > r\} = \mathbb{E}\, e^{-rX}, \quad r \in \mathbb{R}_+.$$

We now show that, in fact, we may take

$$R = Y/X,$$

where Y is independent of X and has the standard exponential distribution: Letting μ denote the distribution of X, for r in $\mathbb{R}_+$,

$$\begin{aligned}
\mathbb{P}\{R > r\} &= \mathbb{P}\{Y > rX\} \\
&= \int_{\mathbb{R}_+} \mu(dx) \int_{\mathbb{R}_+} dy\, e^{-y}\, 1_{(rx,\infty)}(y) \\
&= \int_{\mathbb{R}_+} \mu(dx)\, e^{-rx} = \mathbb{E}\, e^{-rX}
\end{aligned}$$

as was to be shown. In particular, if X has the gamma distribution with shape index a and scale c, then

$$\mathbb{P}\{R > r\} = \mathbb{E}\, e^{-rX} = \left(\frac{c}{c+r}\right)^a, \quad r \in \mathbb{R}_+,$$

according to the Laplace transform computation above in 2.11. Then, R is said to have the *Pareto* distribution with shape index a and scale parameter c. Since $R = Y/X$ and X is "small" in the sense that all its moments are finite, R should be big in the sense that its distribution should have a heavy tail.

Exercises and complements

Some of these are re-statements of results on integrals served up in probabilistic terms. Some are elementary facts that are worth recalling. And some are useful complements. Throughout, X, Y, etc. are random variables.

2.13 *Finiteness.* If $X \geq 0$ and $\mathbb{E}X < \infty$, then $X < \infty$ almost surely. More generally, if X is integrable then it is real-valued almost surely. Show.

2.14 *Moments of positive variables.* If $X \geq 0$, then for every p in $\mathbb{R}_+$,

$$\mathbb{E}X^p = \int_0^\infty dx\; p\; x^{p-1}\; \mathbb{P}\{X > x\}.$$

Show this, using Fubini's theorem with the product measure $\mathbb{P} \times \text{Leb}$, after noting that

$$X^p(\omega) = \int_0^{X(\omega)} dx\; px^{p-1} = \int_0^\infty dx\; px^{p-1}\; 1_{\{X>x\}}(\omega).$$

In particular, if X takes values in $\bar{\mathbb{N}} = \{0, 1, \ldots, +\infty\}$, then

$$\mathbb{E}X = \sum_{n=0}^\infty \mathbb{P}\{X > n\}, \qquad \mathbb{E}X^2 = 2\sum_{n=0}^\infty n\; \mathbb{P}\{X > n\} + \mathbb{E}\; X.$$

2.15 *Optimality of* $\mathbb{E}X$. Define

$$f(a) = \int_\Omega \mathbb{P}(d\omega)\; (X(\omega) - a)^2, \quad a \in \mathbb{R},$$

that is, $f(a)$ is the "weighted sum of errors squared" if X is estimated to be the constant a. Show that f is minimized by $a = \mathbb{E}X$ and that the minimum value is $\text{Var}\; X$.

2.16 *Variance.* Suppose that X is integrable. Show that, for a and b in $\mathbb{R}$,

$$\text{Var}\;(a + bX) \;=\; b^2 \text{Var}\; X.$$

2.17 *Markov's inequality.* For $X \geq 0$,

$$\mathbb{P}\{X > b\} \leq \frac{1}{b}\,\mathbb{E}X$$

for every $b > 0$. Show this by noting that $X \geq b\,1_{\{X>b\}}$.

2.18 *Chebyshev's inequality.* Suppose that X has finite mean. Apply Markov's inequality to $(X - \mathbb{E}X)^2$ to show that

$$\mathbb{P}\{|X - \mathbb{E}X| > \varepsilon\} \leq \frac{1}{\varepsilon^2}\,\mathrm{Var}\,X, \quad \varepsilon > 0.$$

2.19 *Markov's inequality generalized.* Let X be real-valued. Let $f : \mathbb{R} \mapsto \mathbb{R}_+$ be increasing. Show that, for every b in $\mathbb{R}$,

$$\mathbb{P}\{X > b\} \leq \frac{1}{f(b)}\,\mathbb{E}\,f \circ X.$$

2.20 *Jensen's inequality.* Let X have finite mean. Let f be a convex function on $\mathbb{R}$, that is, $f = \sup f_n$ for some sequence of functions f_n having the form $f_n(x) = a_n + b_n x$. Show that

$$\mathbb{E}\,f(X) \;\geq\; f(\mathbb{E}X).$$

2.21 *Gamma distribution.* This is to generalize Example 2.12 slightly by the use of the remark on "scale parameter" in Example 2.10. Let X and Y be independent, let X have distribution $\gamma_{a,c}$ and Y the distribution $\gamma_{b,c}$. Then, show that, $X + Y$ has the distribution $\gamma_{a+b,c}$, and $X/(X+Y)$ has the same old distribution $\beta_{a,b}$, and the two random variables are independent.

2.22 *Gaussian variables.* Show that X has the Gaussian distribution with mean a and variance b if and only if $X = a + \sqrt{b}Z$ for some random variable Z that has the standard Gaussian distribution. Show that

$$\mathbb{E}\,Z = 0, \quad \mathrm{Var}\,Z = 1, \quad \mathbb{E}\,e^{irZ} = e^{-r^2/2},$$
$$\mathbb{E}\,X = a, \quad \mathrm{Var}\,X = b, \quad \mathbb{E}\,e^{irX} = e^{ira - r^2 b/2}.$$

2.23 *Gamma-Gaussian connection.* a) Let Z have the Gaussian distribution with mean 0 and variance b. Show that, then, $X = Z^2$ has the gamma distribution with shape index $a = 1/2$ and scale parameter $c = 1/2b$. Hint: Compute $\mathbb{E}\,f \circ X = \mathbb{E}\,g \circ Z$ with $g(z) = f(z^2)$ and use Theorem 2.4 to identify the result.

b) Let $Z_1, \ldots, Z_n$ be independent standard Gaussian variables. Show that the sum of their squares has the gamma distribution with shape index $n/2$ and scale $1/2$.

2.24 *Uniform distribution.* Let $a < b$ be real numbers. Uniform distribution on (a, b) is the Lebesgue measure on (a, b) normalized to have mass one, that is, $\frac{1}{b-a}$Leb. The standard case is where $a = 0$ and $b = 1$. Since the Lebesgue measure puts no mass at points, the uniform distribution on $[a, b]$ is practically the same as that on (a, b). Let U have the standard uniform distribution on $(0, 1)$; let q be a quantile function. Then $q \circ U$ is a random variable having q as its quantile function.

2.25 *Uniform and exponential.* Let U have the uniform distribution on $(0, 1)$. Let $X = -\frac{1}{c} \log U$. Show that X has the exponential distribution with scale parameter c.

2.26 *Exponential-Gaussian-Uniform.* Let U and V be independent and uniformly distributed on $(0, 1)$. Let $R = \sqrt{-2 \log U}$, so that R^2 has the exponential distribution with scale parameter $1/2$, that is, R^2 has the same distribution as the sum of the squares of two independent standard Gaussian variables.Define

$$X = R \ \cos \ 2\pi V, \qquad Y = R \ \sin \ 2\pi V.$$

Show that X and Y are independent standard Gaussian variables. Show that, conversely, if X and Y are independent standard Gaussian variables, then the polar coordinates R and A of the random point (X, Y) in $\mathbb{R}^2$ are independent, R^2 has the exponential distribution with scale parameter $1/2$, and A has the uniform distribution on $[0, 2\pi]$.

2.27 *Cauchy distribution.* Let X and Y be independent standard Gaussian variables. Show that the distribution μ of $Z = X/Y$ has the form

$$\mu(dz) = dz \ \frac{1}{\pi(1 + z^2)}, \qquad z \in \mathbb{R}.$$

It is called the Cauchy distribution. Note that, if a random variable Z has the Cauchy distribution, then so does $1/Z$. Also, show that, if A has the uniform distribution on $(0, 2\pi)$, then $\tan A$ and $\cot A$ are both Cauchy distributed.

2.28 *Sums and transforms.* Let X and Y be independent positive random variables. Show that the Laplace transform for $X + Y$ is the product of the Laplace transforms for X and Y. Since the Laplace transform of a distribution determines the distribution, this specifies the distribution of $X + Y$, at least in principle. When X and Y are real-valued (instead of being positive), the same statements hold for characteristic functions.

2.29 *Characteristic functions.* Let X and Y be independent gamma distributed random variables with respective shape indices a and b, and the same scale parameter c. Compute the characteristic functions of X, Y, $X+Y$, $X-Y$. Note, in particular, that $X+Y$ has the gamma distribution with shape index $a + b$ and scale c.

2.30 *Gaussian with gamma variance.* Let X and Y be independent, X having the gamma distribution $\gamma_{a,c}$ (with shape index a and scale parameter c), and Y having the standard Gaussian distribution. Recall that $\sqrt{b}Y$ has the Gaussian distribution with mean 0 and variance $b > 0$. We now replace b with X: let $Z = \sqrt{X}\,Y$. Show that

$$\mathbb{E}\, e^{irZ} = \mathbb{E}\, e^{-r^2X/2} = \left(\frac{2c}{2c+r^2}\right)^a, \quad r \in \mathbb{R}.$$

Let U and V be independent with the distribution $\gamma_{a,\sqrt{2c}}$ for both. Show that

$$\mathbb{E}\, e^{ir(U-V)} = \mathbb{E}\, e^{irZ}, \quad r \in \mathbb{R}.$$

Conclude that $\sqrt{X}\,Y$ has the same distribution as $U-V$. (Was the attentive reader able to compute the density in Example 1.16? Can he do it now?)

2.31 *Laplace transforms and finiteness.* Recall that $0 \cdot x = 0$ for all $x \in \mathbb{R}$ and for $x = +\infty$. Thus, if $\hat{\mu}_r = \mathbb{E}\, e^{-rX}$ for some positive random variable X, then $\hat{\mu}_0 = 1$. Show that $r \mapsto \hat{\mu}_r$ is continuous and decreasing on $(0,\infty)$. Its continuity at 0 depends on whether X is almost surely finite: show that

$$\lim_{r\downarrow 0} \hat{\mu}_r = \mathbb{P}\{X < +\infty\}.$$

Hint: For $r > 0$, $e^{-rX} = e^{-rX} 1_{\{X<\infty\}} \nearrow 1_{\{X<\infty\}}$ as $r \downarrow 0$.

2.32 *Laplace transforms and moments.* Let $r \mapsto \hat{\mu}_r$ be the Laplace transform for a positive and almost surely finite random variable X. Use Fubini's theorem for the product measure $\mathbb{P}\times$Leb to show that

$$\int_r^\infty dq\; \mathbb{E}\, Xe^{-qX} = \hat{\mu}_r, \qquad r \in \mathbb{R}_+.$$

This shows, when $\mathbb{E}X$ is finite, that the Laplace transform $\hat{\mu}$ is differentiable on $\mathbb{R}_+^* = (0,\infty)$, and

$$\frac{d}{dr}\hat{\mu}_r = -\mathbb{E}\, Xe^{-rX}, \qquad r \in \mathbb{R}_+^*;$$

in particular, then, the dominated convergence theorem yields

$$\lim_{r\downarrow 0} \frac{d}{dr}\hat{\mu}_r = -\mathbb{E}X.$$

A similar result holds for higher moments: if $\mathbb{E}X^n < \infty$,

$$\lim_{r\downarrow 0} \frac{d^n}{dr^n}\hat{\mu}_r = (-1)^n\, \mathbb{E}X^n.$$

2.33 *Characteristic functions and moments.* Let $\hat{\mu}$ be the characteristic function of a real-valued random variable X. Then, similar to the results of 2.32,

$$\lim_{r\to 0} \frac{d^n}{dr^n}\hat{\mu}_r = i^n\ \mathbb{E}X^n, \qquad n \in \mathbb{N},$$

provided that X^n be integrable, that is, provided that $\mathbb{E}\,|X|^n < \infty$. Generally, the equality above fails when $\mathbb{E}\ |X|^n = \infty$. However, for n even, if the limit on the left is finite, then the equality holds.

2.34 *Uniqueness of distributions and Laplace transforms.* Let X and Y be positive random variables. Show that the following are equivalent:

a) X and Y have the same distribution.
b) $\mathbb{E}\ e^{-rX} = \mathbb{E}\ e^{-rY}$ for every r in $\mathbb{R}_+$.
c) $\mathbb{E}\ f\circ X = \mathbb{E}\ f\circ Y$ for every f bounded continuous.
d) $\mathbb{E}\ f\circ X = \mathbb{E}\ f\circ Y$ for every f bounded Borel.
e) $\mathbb{E}\ f\circ X = \mathbb{E}\ f\circ Y$ for every f positive Borel.

Hint: Show that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a) $\iff$ (e) $\iff$ (d). The difficult parts are (b) $\Rightarrow$ (c) and (c) $\Rightarrow$ (a). For (c) $\Rightarrow$ (a), start by showing that the indicator of an open interval is the limit of an increasing sequence of bounded continuous functions, and use the fact that open intervals form a p-system that generates the Borel σ-algebra on $\mathbb{R}$. For showing (b) $\Rightarrow$ (c), it is useful to recall the following consequence of the Stone-Weierstrass theorem: Let F be the collection of all functions f on $\bar{\mathbb{R}}_+$ having the form

$$f(x) = \sum_{i=1}^{n} c_i e^{-r_i x}$$

for some integer $n \geq 1$, constants $c_1, \ldots, c_n$ in $\mathbb{R}$, and constants $r_1, \ldots, r_n$ in $\mathbb{R}_+$. For every continuous function f on an interval $[a, b]$ of $\mathbb{R}_+$ there exists a sequence in F that converges to f uniformly on $[a, b]$.

2.35 *Uniqueness and characteristic functions.* Let X and Y be real-valued random variables. The statements (a)-(e) in the preceding exercise remain equivalent, except that (b) should be replaced with

b') $\mathbb{E}\ e^{irX} = \mathbb{E}\ e^{irY}$ for every r in $\mathbb{R}$.

2.36 *Random vectors.* Let $X = (X_1, \ldots, X_d)$ be a random variable taking values in $\mathbb{R}^d$, here $d \geq 1$ is an integer. The expected value of X is defined to be the vector

$$\mathbb{E}X = (\mathbb{E}X_1, \ldots, \mathbb{E}X_d).$$

The characteristic function of X is defined to be

$$\mathbb{E}\ e^{ir\cdot X}, \qquad r \in \mathbb{R}^d,$$

where $r \cdot x = r_1 x_1 + \cdots + r_d x_d$, the inner product of r and x. When the components X_i are positive, Laplace transform of the distribution of X is defined similarly: $\mathbb{E}\ e^{-r \cdot X}$, $r \in \mathbb{R}_+^d$. As in the one-dimensional case, the characteristic function determines the distribution of X, and similarly for the Laplace transform. The equivalences in Exercises 2.34 and 2.35 remain true with the obvious modifications: in 2.34(b) and 2.35(b'), r should be in $\mathbb{R}_+^d$ and $\mathbb{R}^d$ respectively, and the functions alluded to should be defined on $\mathbb{R}_+^d$ and $\mathbb{R}^d$ respectively.

2.37 *Covariance.* Let X and Y be real-valued random variables with finite variances. Then, their covariance is defined to be

$$\mathrm{Cov}(X, Y) = \mathbb{E}\ (X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}\ XY - \mathbb{E}X\ \mathbb{E}Y,$$

which is well-defined, is finite, and is bounded in absolute value by $\sqrt{\mathrm{Var}\ X}\ \sqrt{\mathrm{Var}\ Y}$; see Schwartz inequality in the next section. Show that $\mathrm{Var}(X + Y) = \mathrm{Var}\ X + \mathrm{Var}\ Y + 2\mathrm{Cov}(X, Y)$. If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. The converse is generally false.

2.38 *Orthogonality.* Let X and Y be as in 2.37 above. They are said to be orthogonal, or *uncorrelated*, if $\mathbb{E}\ XY = \mathbb{E}X\ \mathbb{E}Y$. So, orthogonality is the same as having vanishing covariance. Show that, if $X_1, \ldots, X_n$ are pairwise orthogonal, that is, X_i and X_j are orthogonal for $i \neq j$, then

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}\ X_1 + \cdots + \mathrm{Var}\ X_n.$$

2.39 *Multi-dimensional Gaussian vectors.* Let X be a d-dimensional random vector; see 2.36 above. It is said to be *Gaussian* if $r \cdot X = r_1 X_1 + \cdots + r_d X_d$ has a Gaussian distribution for every vector r in $\mathbb{R}^d$. It follows that the characteristic function of X has the form

$$\mathbb{E}\ e^{ir \cdot X} = e^{ia(r) - b(r)/2}, \quad r \in \mathbb{R}^d,$$

where $a(r) = \mathbb{E}\ r \cdot X$ and $b(r) = \mathrm{Var}\ r \cdot X$. Let

$$a = (a_1, \ldots, a_d) = (\mathbb{E}\ X_1, \ldots, \mathbb{E}\ X_d) = \mathbb{E}\ X,$$

and let $v = (v_{ij})$ be the $d \times d$ matrix of covariances $v_{ij} = \mathrm{Cov}(X_i, X_j)$. Note that the diagonal entries are variances.

a) Show that $a(r) = a \cdot r$ and $b(r) = r \cdot vr$ where vr is the vector obtained when v is multiplied by the column vector r. Conclude that the distribution of a Gaussian vector X is determined by its mean vector a and covariance matrix v.

b) Show that v is necessarily symmetric and positive definite, that is, $v_{ij} = v_{ji}$ for all i and j, and

$$r \cdot vr = \sum_{i=1}^{d} \sum_{j=1}^{d} r_i v_{ij} r_j \geq 0$$

for every r in $\mathbb{R}^d$.

2.40 *Independence.* Let X be a Gaussian random vector in $\mathbb{R}^d$ with mean vector a and covariance matrix v. Show that X_i and X_j are independent if and only if $v_{ij} = 0$. More generally, if I and J are disjoint subsets of $\{1, \ldots, d\}$, the random vectors $(X_i)_{i \in I}$ and $(X_j)_{j \in J}$ are independent if and only if $v_{ij} = 0$ for every pair (i, j) in $I \times J$. Show.

2.41 *Gaussian distribution.* Let X be a Gaussian vector in $\mathbb{R}^d$ with mean a and covariance matrix v. Then, its characteristic function is given by

$$\mathbb{E}\, e^{ir \cdot X} = e^{ia \cdot r - (r \cdot vr)/2}, \quad r \in \mathbb{R}^d.$$

If v is invertible, that is, if the rank of v is d, the distribution μ of X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, and the corresponding density function is

$$\frac{1}{\sqrt{\det(2\pi v)}} \exp[-\frac{1}{2}(x-a) \cdot v^{-1}(x-a)], \quad x \in \mathbb{R}^d,$$

where v^{-1} is the inverse of v and $\det m$ is the determinant of m; note that $\det(2\pi v) = (2\pi)^d \det v$.

If v is singular, that is, if the rank d' of v is less than d, then at least one entry of the vector X is a linear combination of the other entries. In that case, the distribution μ is no longer absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. Instead, μ puts its mass on some hyperplane of dimension d' in $\mathbb{R}^d$.

2.42 *Continuation.* Let Z_1 and Z_2 be independent standard Gaussian variables (with means 0 and variances 1). Define a random vector X in $\mathbb{R}^3$ by letting $X = cZ$, where

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad c = \begin{bmatrix} 1 & 2 \\ -1 & 3 \\ 4 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

Each X_i is a linear combination of Z_1 and Z_2, therefore every linear combination of X_1, X_2, X_3 is also a linear combination of Z_1, Z_2. So, X is a 3-dimensional Gaussian random vector. Show that its covariance matrix is $v = cc^T$, where c^T is the transpose of c, that is, $v_{ij} = \sum_{k=1}^{2} c_{ik} c_{jk}$. Show that X_3 is a linear combination of X_1 and X_2. Show that Z_1 and Z_2 are linear combinations of X_1 and X_2; find the coefficients involved.

2.43 *Representation of Gaussian vectors.* Every Gaussian random vector X in $\mathbb{R}^d$ has the form

$$X = a + cZ,$$

where a is in $\mathbb{R}^d$, and c is a $d \times d'$ matrix, and Z is a random vector in $\mathbb{R}^{d'}$ whose coordinates are independent one-dimensional standard Gaussian variables. Then, X has mean a and covariance matrix $v = cc^T$.

3 L^p-spaces and Uniform Integrability

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let X be a real-valued random variable. For p in $[1, \infty)$, define

3.1 $$\|X\|_p = (\mathbb{E}\,|X|^p)^{1/p},$$

and for $p = \infty$ let

3.2 $$\|X\|_\infty = \inf\{b \in \mathbb{R}_+ :\ |X| \le b \text{ almost surely}\}.$$

It is easy to see that

3.3 $$\|X\|_p = 0 \quad \Rightarrow \quad X = 0 \text{ almost surely},$$

3.4 $$\|cX\|_p = c\, \|X\|_p\,, \qquad c \ge 0;$$

and it will follow from Theorem 3.6a below with $Y = 1$ that

3.5 $$0 \le \|X\|_p \le \|X\|_q \le +\infty \qquad \text{if } 1 \le p \le q \le +\infty.$$

For each p in $[1, \infty]$, let L^p denote the collection of all real-valued random variables X with $\|X\|_p < \infty$. For p in $[1, \infty)$, X is in L^p if and only if $|X|^p$ is integrable; and X is in L^∞ if and only if X is almost surely bounded. For X in L^p, the number $\|X\|_p$ is called the L^p*-norm* of X; in particular, $\|X\|_\infty$ is called the *essential supremum* of X. Indeed, the properties 3.4 and 3.5 together with Minkowski's inequality below imply that each L^p is a normed vector space provided that we identify X and Y in L^p as one random variable if $X = Y$ almost surely.

Inequalities

The following theorem summarizes the various connections. Its proof will be put after a lemma of independent interest.

3.6 Theorem. *a)* Hölder's inequality*: For* p, q, r *in* $[1, \infty)$ *with* $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$,

$$\|XY\|_r \le \|X\|_p\, \|Y\|_q.$$

In particular, Schwartz's inequality *holds:* $\|XY\|_1 \le \|X\|_2\, \|Y\|_2$.

b) Minkowski's inequality*: For* p *in* $[1, \infty]$,

$$\|X + Y\|_p \le \|X\|_p + \|Y\|_p$$

3.7 Lemma. Jensen's inequality*. Let* D *be a convex domain in* $\mathbb{R}^d$*. Let* $f : D \mapsto \mathbb{R}$ *be continuous and concave. Suppose that* $X_1, \ldots, X_d$ *are integrable random variables and that the vector* $(X_1, \ldots, X_d)$ *belongs to* D *almost surely. Then,*

$$\mathbb{E}\, f(X_1, \ldots, X_d) \le f(\mathbb{E}X_1, \ldots \mathbb{E}X_d).$$

Proof. Since D is convex and $X = (X_1, \ldots, X_d)$ is in D almost surely, the vector $a = \mathbb{E}X = (\mathbb{E}X_1, \ldots \mathbb{E}X_d)$ belongs to D. Let $c_1, \ldots, c_d$ be the direction cosines of a hyperplane in $\mathbb{R}^{d+1}$ lying above the surface f and passing through the point $(a, f(a))$ in $\mathbb{R}^d \times \mathbb{R}$. Then,

$$f(x) \leq f(a) + \sum_1^d (x_i - a_i)\, c_i \qquad x \in D.$$

Replacing x with X and taking expectations yields the desired result. □

In preparation for the proof of Theorem 3.6, we leave it as an exercise to show that, for b in $(0, 1]$,

3.8 $$f(u, v) = u^b\, v^{1-b}, \qquad g(u, v) = (u^b + v^b)^{\frac{1}{b}}$$

define functions that are continuous and concave on $\mathbb{R}^2_+$. Thus, by the preceding lemma,

3.9 $$\mathbb{E}\, U^b V^{1-b} \leq (\mathbb{E}U)^b\, (\mathbb{E}V)^{1-b}, \qquad \mathbb{E}\, (U^b + V^b)^{\frac{1}{b}} \leq [(\mathbb{E}U)^b + (\mathbb{E}V)^b]^{\frac{1}{b}},$$

provided that U and V be positive integrable random variables.

Proof of Theorem 3.6

a) Hölder's inequality. Assume that $\|X\|_p$ and $\|Y\|_q$ are finite; otherwise, there is nothing to prove. When $p = \infty$, we have $|XY| \leq \|X\|_p\, |Y|$ almost surely, and hence the inequality is immediate; similarly for $q = \infty$. Assuming that p and q are both finite, the inequality desired follows from the first inequality in 3.9 with $b = \frac{r}{p}$, $U = |X|^p$, $V = |Y|^q$.

b) Minkowski's inequality. Again, assume that $\|X\|_p$ and $\|Y\|_p$ are finite. If $p = \infty$, the inequality is immediate from the definition 3.2 (and in fact, becomes an equality). For p in $[1, \infty)$, the inequality follows from the second inequality in 3.9 with $b = \frac{1}{p}$, $U = |X|^p$, $V = |Y|^p$. □

Uniform integrability

This concept plays an important role in martingale theory and in the convergence of sequences in the space L^1. We start by illustrating the issue involved in the simplest setting.

3.10 Lemma. *Let X be a real-valued random variable. Then, X is integrable if and only if*

3.11 $$\lim_{b \to \infty} \mathbb{E}\, |X|\, 1_{\{|X| > b\}} = 0.$$

Proof. Let Z_b denote the variable inside the expectation in 3.11. Note that it is dominated by $|X|$ and goes to 0 as $b \to \infty$. Thus, if X is integrable, the dominated convergence yields that $\lim_{b\to\infty} \mathbb{E}\ Z_b = 0$, which is exactly 3.11. Conversely, if 3.11 holds, then we can choose b large enough to have $\mathbb{E}\ Z_b \leq 1$, and the inequality $|X| \leq b + Z_b$ shows that $\mathbb{E}\ |X| \leq b + 1 < \infty$. □

For a collection of random variables X, the uniform integrability of the collection has to do with the possibility of taking the limit in 3.11 uniformly in X:

3.12 DEFINITION. *A collection $\mathcal{K}$ of real-valued random variables is said to be* uniformly integrable *if*

$$k(b) = \sup_{X\in\mathcal{K}} \mathbb{E}\ |X|\ 1_{\{|X|>b\}}$$

goes to 0 as $b \to \infty$.

3.13 REMARKS. a) If $\mathcal{K}$ is finite and each X in it is integrable, then $\mathcal{K}$ is uniformly integrable. For, then, the limit over b, of $k(b)$, can be passed inside the supremum, and Lemma 3.10 does the rest.

b) If $\mathcal{K}$ is dominated by an integrable random variable Z, then it is uniformly integrable. Because, then, $|X| \leq Z$ for every X in $\mathcal{K}$, which yields $k(b) \leq \mathbb{E}\ Z\ 1_{\{Z>b\}}$, and that last expectation goes to 0 by Lemma 3.10 applied to Z.

c) Uniform integrability implies L^1-boundedness, that is, if $\mathcal{K}$ is uniformly integrable then $\mathcal{K} \subset L^1$ and

$$k(0) = \sup_{\mathcal{K}} \mathbb{E}\ |X| < \infty.$$

To see this, note that $\mathbb{E}\ |X| \leq b + k(b)$ for all X and use the uniform integrability of $\mathcal{K}$ to choose a finite number b such that $k(b) \leq 1$.

d) But L^1-boundedness is insufficient for uniform integrability. Here is a sequence $\mathcal{K} = \{X_n : n \geq 1\}$ that is L^1-bounded but not uniformly integrable. Suppose that $\Omega = (0,1)$ with its Borel σ-algebra for events and the Lebesgue measure as $\mathbb{P}$. Let $X_n(\omega)$ be equal to n if $\omega \leq 1/n$ and to 0 otherwise. Then, $\mathbb{E}\ X_n = 1$ for all n, that is, $\mathcal{K}$ is L^1-bounded. But $k(b) = 1$ for all b, since $\mathbb{E}\ X_n\ 1_{\{X_n>b\}} = \mathbb{E}\ X_n = 1$ for $n > b$.

e) However, if $\mathcal{K}$ is L^p-bounded for some $p > 1$ then it is uniformly integrable. This will be shown below: see Proposition 3.17 and take $f(x) = x^p$.

The following ε-δ characterization is the main result on uniform integrability: over every small set, the integrals of the X are uniformly small.

3.14 THEOREM. *The collection $\mathcal{K}$ is uniformly integrable if and only if it is L^1-bounded and for every $\varepsilon > 0$ there is $\delta > 0$ such that, for every event H,*

3.15 $$\mathbb{P}(H) \leq \delta \Rightarrow \sup_{X\in\mathcal{K}} \mathbb{E}\ |X|\ 1_H \leq \varepsilon.$$

Proof. We assume that all X are positive; this amounts to working with $|X|$ throughout. Since $X\ 1_H \le b\ 1_H + X\ 1_{\{X>b\}}$ for every event H and every b in $\mathbb{R}_+$,

3.16 $$\sup_{X\in\mathcal{K}} \mathbb{E}\ X\ 1_H \le b\mathbb{P}(H) + k(b), \quad b\in\mathbb{R}_+.$$

Suppose that $\mathcal{K}$ is uniformly integrable. Then, it is L^1-bounded by Remark 3.13c. Also, since $k(b) \to 0$, for every $\varepsilon > 0$ there is $b < \infty$ such that $k(b) \le \varepsilon/2$, and setting $\delta = \varepsilon/2b$ we see that 3.15 holds in view of 3.16.

Conversely, suppose that $\mathcal{K}$ is L^1-bounded and that for every $\varepsilon > 0$ there is $\delta > 0$ such that 3.15 holds for all events H. Then, Markov's inequality 2.17 yields

$$\sup_{X\in\mathcal{K}} \mathbb{P}\{X>b\} \le \frac{1}{b}\sup_{X\in\mathcal{K}} \mathbb{E}\ X = \frac{1}{b}k(0),$$

which shows the existence of b such that $\mathbb{P}\{X>b\} \le \delta$ for all X, and, then, for that b we have $k(b) \le \varepsilon$ in view of 3.15 used with $H = \{X > b\}$. In other words, for every $\varepsilon > 0$ there is $b < \infty$ such that $k(b) \le \varepsilon$, which is the definition of uniform integrability. □

The following proposition is very useful for showing uniform integrability. In particular, as remarked earlier, it shows that L^p-boundedness for some $p > 1$ implies uniform integrability.

3.17 PROPOSITION. *Suppose that there is a positive Borel function f on $\mathbb{R}_+$ such that* $\lim_{x\to\infty} f(x)/x = \infty$ *and*

3.18 $$\sup_{X\in\mathcal{K}} \mathbb{E}\ f\circ|X| < \infty.$$

Then, $\mathcal{K}$ is uniformly integrable.

Proof. We may and do assume that all X are positive. Also, by replacing f with $f \vee 1$ if necessary, we assume that $f \ge 1$ in addition to satisfying the stated conditions. Let $g(x) = x/f(x)$ and note that

$$X\ 1_{\{X>b\}} = f\circ X\ g\circ X\ 1_{\{X>b\}} \le f\circ X\ \sup_{x>b} g(x).$$

This shows that, with c denoting the supremum in 3.18,

$$k(b) \le c\ \sup_{x>b} g(x),$$

and the right side goes to 0 as $b\to\infty$ since $g(x)\to 0$ as $x\to+\infty$. □

We supplement the preceding proposition by a converse and give another characterization.

3.19 Theorem. *The following are equivalent:*

a) $\mathcal{K}$ *is uniformly integrable.*
b) $h(b) = \sup_{\mathcal{K}} \int_b^\infty dy\ \mathbb{P}\{|X| > y\} \to 0$ *as* $b \to \infty$.
c) $\sup_{\mathcal{K}} \mathbb{E}\ f \circ |X| < \infty$ *for some increasing convex function* f *on* $\mathbb{R}_+$ *with* $\lim_{x\to\infty} f(x)/x = +\infty$.

Proof. The preceding proposition shows that (c) ⇒ (a). We now show that (a) ⇒ (b) ⇒ (c), again assuming, as we may, that all the X in $\mathcal{K}$ are positive.

Assume (a). For every X in $\mathcal{K}$,

$$\begin{aligned}\mathbb{E}\ X\ 1_{\{X>b\}} &= \int_0^\infty dy\ \mathbb{P}\{X\ 1_{\{X>b\}} > y\} \\ &= \int_0^\infty dy\ \mathbb{P}\{X > b \vee y\}\ \geq\ \int_b^\infty dy\ \mathbb{P}\{X > y\}.\end{aligned}$$

Thus, $k(b) \geq h(b)$ for every b, and the uniform integrability of $\mathcal{K}$ means that $k(b) \to 0$ as $b \to \infty$. Hence, (a) ⇒ (b).

Assume (b). Since $h(b) \to 0$ as $b \to \infty$, we can pick $0 = b_0 < b_1 < b_2 < \cdots$ increasing to $+\infty$ such that

$$h(b_n) \leq h(0)/2^n, \quad n \in \mathbb{N};$$

note that $h(0)$ is finite since $h(0) \leq b + h(b)$ and $h(b)$ can be made as small as desired. Define

$$g(x) = \sum_{n=0}^\infty 1_{[b_n,\infty)}(x), \quad f(x) = \int_0^x dy\ g(y), \quad x \in \mathbb{R}_+.$$

Note that $g \geq 1$ and is increasing toward $+\infty$, which implies that f is increasing and convex and $\lim_{x\to\infty} f(x)/x = +\infty$. Now,

$$\begin{aligned}\mathbb{E}\ f\circ X &= \mathbb{E} \int_0^X dy\ g(y) \\ &= \sum_{n=0}^\infty \mathbb{E} \int_{b_n}^\infty dy\ 1_{\{X>y\}}\ \leq\ \sum_{n=0}^\infty h(b_n)\ \leq\ 2h(0)\ <\ \infty.\end{aligned}$$

This being true for all X in $\mathcal{K}$, we see that (b) ⇒ (c). □

Exercises and complements

3.20 *Concavity.* Show that the functions f and g defined by 3.8 are continuous and concave. Hint: Note that $f(cu, cv) = c\,f(u, v)$ for every $c > 0$; conclude that it is sufficient to show that $x \mapsto f(x, 1-x)$ from $[0, 1]$ into $\mathbb{R}_+$ is continuous and concave; and show the latter by noting that the second derivative is negative. Similarly for g.

3.21 *Continuity of the norms.* Fix a random variable X. Define $f(p) = \|X\|_p$ for p in $[1, \infty]$. Show that the function f is continuous except possibly at one point p_0, where p_0 is such that

$$f(p) < \infty \text{ for } p < p_0\,, \qquad f(p) = +\infty \text{ for } p > p_0\,,$$

and f is left-continuous at p_0.

3.22 *Integrals over small sets.* Let X be positive and integrable. Let (H_n) be a sequence of events. If $\mathbb{P}(H_n) \to 0$, then $\mathbb{E}X1_{H_n} \to 0$. Show.

3.23 *Uniform integrability.* Let (X_i) and (Y_i) be uniformly integrable. Show that, then,

a) $(X_i \vee Y_i)$ is uniformly integrable,
b) $(X_i + Y_i)$ is uniformly integrable.

3.24 *Comparisons.* If $|X_i| \le |Y_i|$ for each i, and (Y_i) is uniformly integrable, then so is (X_i). Show.

4 Information and Determinability

This section is on σ-algebras generated by random variables and measurability with respect to them. Also, we shall argue that such a σ-algebra should be thought as a body of information, and measurability with respect to it should be equated to being determined by that information. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a probability space.

Sigma-algebras generated by random variables

Let X be a random variable taking values in some measurable space $(E, \mathcal{E})$. Then,

4.1 $$\sigma X = X^{-1}\mathcal{E} = \{X^{-1}A : \quad A \in \mathcal{E}\}$$

is a σ-algebra (and is a subset of $\mathcal{H}$ by the definition of random variables). It is called the *σ-algebra generated by X*, and the notation σX is preferred over the others. Clearly, σX is the smallest σ-algebra $\mathcal{G}$ on Ω such that X is measurable with respect to $\mathcal{G}$ and $\mathcal{E}$; see Exercise I.2.20.

Let T be an arbitrary index set, countable or uncountable. For each t in T let X_t be a random variable taking values in some measurable space $(E_t, \mathcal{E}_t)$. Then,

4.2 $$\sigma\{X_t : \quad t \in T\} = \bigvee_{t \in T} \sigma X_t$$

denotes the σ-algebra on Ω generated by the union of the σ-algebras σX_t, $t \in T$; see Exercise I.1.18. It is called the *σ-algebra generated* by the collection $\{X_t : \ t \in T\}$. It is the smallest σ-algebra $\mathcal{G}$ on Ω such that, for every t in T, the random variable X_t is measurable with respect to $\mathcal{G}$ and $\mathcal{E}_t$; obviously, $\mathcal{G} \subset \mathcal{H}$.

In view of Proposition I.6.27, we may regard the collection $\{X_t : t \in T\}$ as one random variable X taking values in the product space $(E, \mathcal{E}) = \otimes_{t\in T}(E_t, \mathcal{E}_t)$ by defining $X(\omega)$ to be the point $(X_t(\omega))_{t\in T}$ in the "function" space E for each ω. Conversely, if X is a random variable taking values in the product space $(E, \mathcal{E})$, we denote by $X_t(\omega)$ the value of the function $X(\omega)$ at the point t in T; the resulting mapping $\omega \mapsto X_t(\omega)$ is a random variable with values in $(E_t, \mathcal{E}_t)$ and is called the *t-coordinate* of X. It will be convenient to write $X = (X_t)_{t\in T}$ and consider X both as the E-valued random variable and as the collection of random variables X_t, $t \in T$. This causes no ambiguity for σX:

4.3 PROPOSITION. *If $X = (X_t)_{t\in T}$, then $\sigma X = \sigma\{X_t : t \in T\}$.*

Proof. Proof is immediate from that of Proposition I.6.27. Let $\mathcal{H}$ there be σX to conclude that $\sigma X \supset \sigma\{X_t : t \in T\}$, and then let $\mathcal{H}$ be $\sigma\{X_t : t \in T\}$ to conclude that $\sigma\{X_t : t \in T\} \supset \sigma X$. □

Measurability

The following theorem is to characterize the σ-algebra σX. It shows that a random variable is σX-measurable if and only if it is a deterministic measurable function of X. In other words, with the usual identification of a σ-algebra with the collection of all numerical mappings that are measurable relative to it, the collection σX of random variables is exactly the set of all measurable functions of X.

4.4 THEOREM. *Let X be a random variable taking values in some measurable space $(E, \mathcal{E})$. A mapping $V : \Omega \to \bar{\mathbb{R}}$ belongs to σX if and only if*

$$V = f \circ X$$

for some deterministic function f in $\mathcal{E}$.

Proof. *Sufficiency.* Since X is measurable with respect to σX and $\mathcal{E}$, and since measurable functions of measurable functions are measurable, every V having the form $f \circ X$ for some f in $\mathcal{E}$ is σX-measurable.

Necessity. Let $\mathcal{M}$ be the collection of all V having the form $V = f \circ X$ for some f in $\mathcal{E}$. We shall use the monotone class theorem I.2.19 to show that $\mathcal{M} \supset \sigma X$, which is the desired result. We start by showing that $\mathcal{M}$ is a monotone class of functions on Ω.

i) $1 \in \mathcal{M}$ since $1 = f \circ X$ with $f(x) = 1$ for all x in E.

ii) Let U and V be bounded and in $\mathcal{M}$, and let a and b be in $\mathbb{R}$. Then, $U = f \circ X$ and $V = g \circ X$ for some f and g in $\mathcal{E}$, and thus, $aU + bV = h \circ X$ with $h = af + bg$. Since $h \in \mathcal{E}$, it follows that $aU + bV \in \mathcal{M}$.

iii) Let $(V_n) \subset \mathcal{M}_+$ and $V_n \nearrow V$. For each n, there is f_n in $\mathcal{E}$ such that $V_n = f_n \circ X$. Then, $f = \sup f_n$ belongs to $\mathcal{E}$ and since $V_n \nearrow V$,

$$V(\omega) = \sup_n V_n(\omega) = \sup_n f_n(X(\omega)) = f(X(\omega)) , \qquad \omega \in \Omega,$$

which shows that $V \in \mathcal{M}$.

Furthermore, $\mathcal{M}$ includes every indicator variable in σX: if $H \subset \Omega$ is in σX, then $H = X^{-1}A$ for some set A in $\mathcal{E}$, and $1_H = 1_A \circ X \in \mathcal{M}$. Therefore, by the monotone class theorem, $\mathcal{M}$ contains all positive random variables in σX.

Finally, let V in σX be arbitary. Then, $V^+ \in \sigma X$ and is positive, and hence, $V^+ = g \circ X$ for some g in $\mathcal{E}$; similarly, $V^- = h \circ X$ for some h in $\mathcal{E}$. Thus, $V = V^+ - V^- = f \circ X$, where

$$f(x) = \begin{cases} g(x) - h(x) & \text{if } g(x) \wedge h(x) = 0, \\ 0 & \text{otherwise.} \end{cases}$$

This completes the proof since $f \in \mathcal{E}$. □

4.5 Corollary. *For each n in $\mathbb{N}^*$, let X_n be a random variable taking values in some measurable space $(E_n, \mathcal{E}_n)$. A mapping $V : \Omega \mapsto \bar{\mathbb{R}}$ belongs to $\sigma\{X_n : n \in \mathbb{N}^*\}$ if and only if*

$$V = f(X_1, X_2, \ldots)$$

for some f in $\otimes_n \mathcal{E}_n$.

Proof. Proof is immediate from the preceding theorem upon putting $X = (X_1, X_2, \ldots)$ and using Proposition 4.3. □

The preceding corollary can be generalized to uncountable collections $\{X_t : t \in T\}$ by using the same device of regarding the collection as one random variable. In fact, there is a certain amount of simplification, reflecting the fact that uncountable products of σ-algebras $\mathcal{E}_t,\ t \in T$, are in fact generated by the finite-dimensional rectangles.

4.6 Proposition. *Let T be arbitrary. For each t in T, let X_t be a random variable taking values in some measurable space $(E_t, \mathcal{E}_t)$. Then, $V : \Omega \mapsto \bar{\mathbb{R}}$ belongs to $\sigma\{X_t : t \in T\}$ if and only if there exists a sequence (t_n) in T and a function f in $\otimes_n \mathcal{E}_{t_n}$ such that*

4.7 $$V = f(X_{t_1}, X_{t_2}, \ldots).$$

Proof. Sufficiency of the condition is trivial: if V has the form 4.7, then $V \in \sigma\{X_{t_n} : n \geq 1\} = \hat{\mathcal{G}}$ by the corollary above, and $\hat{\mathcal{G}} \subset \mathcal{G} = \sigma\{X_t : t \in T\}$ obviously.

To show the necessity, we use the monotone class theorem I.2.19 together with Proposition 4.3. To that end, let $\mathcal{M}$ be the collection of all V having the form 4.7 for some sequence (t_n) in T and some f in $\otimes\mathcal{E}_{t_n}$. It is easy to check that $\mathcal{M}$ is a monotone class. We shall show that $\mathcal{M}$ includes the indicators of a p-system $\mathcal{G}_0$ that generates $\mathcal{G}$. Then, by the monotone class theorem, $\mathcal{M}$ includes all positive V in $\mathcal{G}$ and, therefore, all V in $\mathcal{G}$ since $V = V^+ - V^-$ is obviously in $\mathcal{M}$ if V^+ and V^- are in $\mathcal{M}$. Hence, $\mathcal{M} \supset \mathcal{G}$ as desired.

By Proposition 4.3, $\mathcal{G} = \sigma X$ where $X = (X_t)_{t\in T}$ takes values in $(E, \mathcal{E})= \bigotimes(E_t, \mathcal{E}_t)$. Recall that $\mathcal{E}$ is generated by the p-system of all finite-dimensional measurable rectangles. Therefore, the inverse images $X^{-1}A$ of those rectangles A form a p-system $\mathcal{G}_0$ that generates $\mathcal{G}$. Thus, to complete the proof, it is sufficient to show that the indicator of $X^{-1}A = \{X \in A\}$ belongs to $\mathcal{M}$ for every such rectangle A.

Let A be such a rectangle, that is, $A = \times_t A_t$ with $A_t = E_t$ for all t outside a finite subset S of T and $A_t \in \mathcal{E}_t$ for every t in S (and therefore for all t in T). Then,

$$1_{\{X\in A\}} = 1_A \circ X = \prod_{t\in S} 1_{A_t} \circ X_t,$$

which has the form 4.7, that is, belongs to $\mathcal{M}$. □

Heuristics

Our aim is to use the foregoing to argue that a σ-algebra on Ω is the mathematically precise equivalent of the everyday term “information”. And, random quantities that are determined by that information are precisely the random variables that are measurable with respect to that σ-algebra.

To fix the ideas, consider a random experiment that consists of a sequence of trials, at each of which there are five possible results labeled a, b, c, d, e. Each possible outcome of this experiment can be represented by a sequence $\omega = (\omega_1, \omega_2, \ldots)$ where $\omega_n \in E = \{a, \ldots, e\}$ for each n. The sample space Ω, then, consists of all such sequences ω. We define $X_1, X_2, \ldots$ to be the coordinate variables, that is $X_n(\omega) = \omega_n$ for every n and outcome ω. We let $\mathcal{H}$ be the σ-algebra generated by $\{X_n : n \in \mathbb{N}^*\}$. The probability $\mathbb{P}$ is unimportant for our current purposes and we leave it unspecified.

Consider the information we shall have about this experiment at the end of the third trial. At that time, whatever the possible outcome ω may be, we shall know $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, and nothing more. In other words, the information we shall have will specify the results $\omega_1, \omega_2, \omega_3$ but nothing more. Thus, the information we shall have will determine the values $V(\omega)$, for every possible ω, provided that the dependence of $V(\omega)$ on ω is through $\omega_1, \omega_2, \omega_3$, that is, provided that $V = f(X_1, X_2, X_3)$ for some deterministic function f on $E\times E\times E$. Based on these arguments, we equate “the information available at the end of third trial” to the σ-algebra $\mathcal{G}$ consisting of all such numerical random variables whose values are determined by that body of information.

In this case, the information $\mathcal{G}$ is generated by $\{X_1, X_2, X_3\}$ in the sense that knowing X_1, X_2, X_3 is equivalent to knowing the information $\mathcal{G}$.

Going back to an arbitrary probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{H}$, we may heuristically equate $\mathcal{G}$ to the information available to someone who is able to tell the value $V(\omega)$ for every possible ω and every random variable V that is $\mathcal{G}$-measurable. Incidentally, this gives a mathematical definition for the imprecise everyday term "information".

Often, there are simpler ways of characterizing the information $\mathcal{G}$. If there is a random variable X such that the knowledge of its value is sufficient to determine the values of all the V in $\mathcal{G}$, then we say that X generates the information $\mathcal{G}$ and write $\mathcal{G} = \sigma X$. This is the heuristic content of the definition of σX.

Of course, embedded in the heuristics is the basic theorem of this section, Theorem 4.4, which now becomes obvious: if the information $\mathcal{G}$ consists of the knowledge of X, then $\mathcal{G}$ determines exactly those variables V that are deterministic functions of X. Another result that becomes obvious is Proposition 4.3: in the setting of it, since knowing X is the same as knowing X_t for all t in T, the information generated by X is the same as the information generated by $X_t,\ t \in T$.

Filtrations

Continuing with the heuristics, suppose that we are interested in a random experiment taking place over an infinite expanse of time. Let $T = \mathbb{R}_+$ or $T = \mathbb{N}$ be the time set. For each time t, let $\mathcal{F}_t$ be the information gathered during $[0, t]$ by an observer of the experiment. For $s < t$, we must have $\mathcal{F}_s \subset \mathcal{F}_t$. The family $\mathcal{F} = \{\mathcal{F}_t :\ t \in T\}$, then, depicts the flow of information as the experiment progresses over time. The following definition formalizes this concept.

4.8 Definition. *Let T be a subset of $\mathbb{R}$. For each t in T, let $\mathcal{F}_t$ be a sub-σ-algebra of $\mathcal{H}$. The family $\mathcal{F} = \{\mathcal{F}_t :\ t \in T\}$ is called a* filtration *provided that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$.*

In other words, a filtration is an increasing family of sub-σ-algebras of $\mathcal{H}$. The simplest examples are the filtrations generated by stochastic processes: If $X = \{X_t :\ t \in T\}$ is a stochastic process, then putting $\mathcal{F}_t = \sigma\{X_s :\ s \le t,\ s \in T\}$ yields a filtration $\mathcal{F} = \{\mathcal{F}_t :\ t \in T\}$. The reader is invited to ponder the meaning of the next proposition for such a filtration. Of course, the aim is to approximate eternal variables by random variables that become known in finite time.

4.9 Proposition. *Let $\mathcal{F} = \{\mathcal{F}_n :\ n \in \mathbb{N}\}$ be a filtration and put $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$. For each bounded random variable V in $\mathcal{F}_\infty$ there are bounded variables V_n in $\mathcal{F}_n$, $n \in \mathbb{N}$, such that*

$$\lim_n \ \mathbb{E}\, |V_n - V| \ = \ 0.$$

Remark. Note that $\mathbb{E}|V_n - V| = \|V_n - V\|_1$ in the notation of section 3; thus, the approximation here is in the sense of L^1-space. Also, we may add to the conclusion that $\mathbb{E}V_n \to \mathbb{E}V$; this follows from the observation that $|\mathbb{E}V_n - \mathbb{E}V| \le \mathbb{E}|V_n - V|$.

Proof. Let $\mathcal{C} = \bigcup_n \mathcal{F}_n$. By definition, $\mathcal{F}_\infty = \sigma\mathcal{C}$. Obviously $\mathcal{C}$ is a p-system. To complete the proof via the monotone class theorem, we start by letting $\mathcal{M}_b$ be the collection of all bounded variables in $\mathcal{F}_\infty$ having the approximation property described. It is easy to see that $\mathcal{M}_b$ includes constants and is a vector space over $\mathbb{R}$ and includes the indicators of events in $\mathcal{C}$. Thus, $\mathcal{M}_b$ will include all bounded V in $\mathcal{F}_\infty$ once we check the remaining monotonicity condition.

Let $(U_k) \subset \mathcal{M}_b$ be positive and increasing to a bounded variable V in $\mathcal{F}_\infty$. Then, for each $k \ge 1$ there are $U_{k,n}$ in $\mathcal{F}_n$, $n \in \mathbb{N}$, such that $\mathbb{E}|U_{k,n} - U_k| \to 0$ as $n \to \infty$. Put $n_0 = 0$, and for each $k \ge 1$ choose $n_k > n_{k-1}$ such that $\hat{U}_k = U_{k,n_k}$ satisfies

$$\mathbb{E}|\hat{U}_k - U_k| < \frac{1}{k}.$$

Moreover, since (U_k) is bounded and converges to V, the bounded convergence implies that $\mathbb{E}|U_k - V| \to 0$. Hence,

4.10 $$\mathbb{E}\,|\hat{U}_k - V| \;\le\; \mathbb{E}\,|\hat{U}_k - U_k| \;+\; \mathbb{E}\,|U_k - V| \;\to\; 0$$

as $k \to \infty$. With $n_0 = 0$ choose $V_0 = 0$ and put $V_n = \hat{U}_k$ for all integers n in $(n_k, n_{k+1}]$; then, $V_n \in \mathcal{F}_{n_k} \subset \mathcal{F}_n$, and $\mathbb{E}|V_n - V| \to 0$ as $n \to \infty$ in view of 4.10. This is what we need to show that $V \in \mathcal{M}_b$. □

In the preceding proposition, the V_n are shown to exist but are unspecified. A very specific version will appear later employing totally new tools; see the martingale convergence theorems of Chapter V and, in particular, Corollary V.3.30 there.

Exercises and complements

4.11 *p-systems for σX.* Let T be an arbitrary index set. Let $X = (X_t)_{t\in T}$, where X_t takes values in $(E_t, \mathcal{E}_t)$ for each t in T. For each t, let $\mathcal{C}_t$ be a p-system that generates $\mathcal{E}_t$. Let $\mathcal{G}_0$ be the collection of all $G \subset \Omega$ having the form

$$G = \bigcap_{t\in S} \{X_t \in A_t\}$$

for some finite $S \subset T$ and A_t in $\mathcal{C}_t$ for every t in S. Show that $\mathcal{G}_0$ is a p-system that generates $\mathcal{G} = \sigma X$.

4.12 *Monotone class theorem.* This is a generalization of the monotone class theorem I.2.19. We keep the setting and notations of the preceding exercise.

Let $\mathcal{M}$ be a monotone class of mappings from Ω into $\bar{\mathbb{R}}$. Suppose that $\mathcal{M}$ includes every $V : \Omega \mapsto [0,1]$ having the form

$$V = \prod_{t \in S} 1_{A_t} \circ X_t \ , \qquad S \text{ finite, } A_t \in \mathcal{C}_t \text{ for every } t \text{ in } S.$$

Then, every positive V in σX belongs to $\mathcal{M}$. Prove.

4.13 *Special case.* In the setting of the exercises above, suppose $E_t = \mathbb{R}$ and $\mathcal{E}_t = \mathcal{B}_{\mathbb{R}}$ for all t. Let $\mathcal{M}$ be a monotone class of mappings from Ω into $\bar{\mathbb{R}}$. Suppose that $\mathcal{M}$ includes every V of the form

$$V = f_1 \circ X_{t_1} \cdots f_n \circ X_{t_n}$$

with $n \geq 1$ and $t_1, \ldots, t_n$ in T and $f_1, \ldots, f_n$ bounded continuous functions from $\mathbb{R}$ into $\mathbb{R}$. Then, $\mathcal{M}$ contains all positive V in σX. Prove. Hint: Start by showing that, if A is an open interval of $\mathbb{R}$, then 1_A is the limit of an increasing sequence of bounded continuous functions.

4.14 *Determinability.* If X and Y are random variables taking values in $(E, \mathcal{E})$ and $(D, \mathcal{D})$, then we say that X *determines* Y if $Y = f \circ X$ for some $f : E \mapsto D$ measurable with respect to $\mathcal{E}$ and $\mathcal{D}$. Then, $\sigma X \supset \sigma Y$ obviously. Heuristically, X determines Y if knowing $X(\omega)$ is sufficient for knowing $Y(\omega)$, this being true for every possibility ω. To illustrate the notion in a simple setting, let T be a positive random variable and define a stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$ by setting, for each ω

$$X_t(\omega) = \begin{cases} 0 & \text{if } t < T(\omega), \\ 1 & \text{if } t \geq T(\omega). \end{cases}$$

Show that X and T determine each other. If T represents the time of failure for a device, then X is the process that indicates whether the device has failed or not. That X and T determine each other is intuitively obvious, but the measurability issues cannot be ignored altogether.

4.15 *Warning.* A slight change in the preceding exercise shows that one must guard against raw intuition. Let T have a distribution that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$; in fact, all we need is that $\mathbb{P}\{T = t\} = 0$ for every t in $\mathbb{R}_+$. Define

$$X_t(\omega) = \begin{cases} 1 & \text{if } t = T(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

Show that, for each t in $\mathbb{R}_+$, the random variable X_t is determined by T. But, contrary to raw intuition, T is not determined by $X = (X_t)_{t \in \mathbb{R}_+}$. Show this by following the steps below:

a) For each t, we have $X_t = 0$ almost surely. Therefore, for every sequence (t_n) in $\mathbb{R}_+$, $X_{t_1} = X_{t_2} = \ldots = 0$ almost surely.

b) If $V \in \sigma X$, then $V = c$ almost surely for some constant c. It follows that T is not in σX.

4.16 *Arrival processes.* Let $T = (T_1, T_2, \ldots)$ be an increasing sequence of $\mathbb{R}_+$-valued variables. Define a stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$ with state space $\mathbb{N}$ by

$$X_t = \sum_{n=1}^{\infty} 1_{(0,t]} \circ T_n \ , \qquad t \in \mathbb{R}_+ .$$

Show that X and T determine each other. If T_n represents the n-th arrival time at a store, then X_t is the number of customers who arrived during $(0, t]$. So, X and T are the same phenomena viewed from different angles.

5 Independence

This section is about independence, a truly probabilistic concept. For random variables, the concept reduces to the earlier definition: they are independent if and only if their joint distribution is the product of their marginal distributions.

Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a probability space. As usual, if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{H}$, we regard it both as a collection of events and as the collection of all numerical random variables that are measurable with respect to it. Recall that σX is the σ-algebra on Ω generated by X, and X here can be a random variable or a collection of random variables. Finally, we write $\mathcal{F}_I$ for $\bigvee_{i \in I} \mathcal{F}_i$ as in I.1.8 and refer to it as the σ-algebra generated by the collection of σ-algebras $\mathcal{F}_i$, $i \in I$.

Definitions

For a fixed integer $n \geq 2$, let $\mathcal{F}_1, \ldots \mathcal{F}_n$ be sub-σ-algebras of $\mathcal{H}$. Then, $\{\mathcal{F}_1, \ldots \mathcal{F}_n\}$ is called an *independency* if

5.1 $$\mathbb{E}\, V_1 \cdots V_n = \mathbb{E} V_1 \cdots \mathbb{E} V_n$$

for all positive random variables $V_1, \ldots, V_n$ in $\mathcal{F}_1, \ldots, \mathcal{F}_n$ respectively. The term "independency" is meant to suggest a realm governed by the independence of its constituents.

Let T be an arbitrary index set. Let $\mathcal{F}_t$ be a sub-σ-algebra of $\mathcal{H}$ for each t in T. The collection $\{\mathcal{F}_t : \ t \in T\}$ is called an *independency* if its every finite subset is an independency.

In general, elements of an independency are said to be *independent*, or *mutually independent* if emphasis is needed. In loose language, given some objects, the objects are said to be *independent* if the σ-algebras generated by those objects are independent. The objects themselves can be events, random variables, collections of random variables, σ-algebras on Ω, or collections of such, and so on, and they might be mixed. For example, a random variable

X and a stochastic process $\{Y_t : t \in T\}$ and a collection $\{\mathcal{F}_i : i \in I\}$ of σ-algebras on Ω are said to be *independent* if

$$\mathcal{G}_1 = \sigma X , \quad \mathcal{G}_2 = \sigma\{Y_t : t \in T\} , \quad \mathcal{G}_3 = \mathcal{F}_I = \bigvee_{i \in I} \mathcal{F}_i$$

are independent, that is, if $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\}$ is an independency.

Independence of σ-algebras

Since a collection of sub-σ-algebras of $\mathcal{H}$ is an independency if and only if its every finite subset is an independency, we concentrate on the independence of a finite number of sub-σ-algebras of $\mathcal{H}$. We start with a test for independence.

5.2 PROPOSITION. *Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ be sub-σ-algebras of $\mathcal{H}$, $n \geq 2$. For each $i \leq n$, let $\mathcal{C}_i$ be a p-system that generates $\mathcal{F}_i$. Then, $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if and only if*

5.3 $$\mathbb{P}(H_1 \cap \cdots \cap H_n) = \mathbb{P}(H_1) \cdots \mathbb{P}(H_n)$$

for all H_i in $\bar{\mathcal{C}}_i = \mathcal{C}_i \cup \{\Omega\}$, $i = 1, \ldots, n$.

Proof. Necessity is obvious: take the V_i in 5.1 to be the indicators of the events H_i. To show the sufficiency part, assume 5.3 for H_i in $\bar{\mathcal{C}}_i$, $i = 1, \ldots, n$. Fix $H_2, \ldots, H_n$ in $\bar{\mathcal{C}}_2, \ldots, \bar{\mathcal{C}}_n$ respectively, and let $\mathcal{D}$ be the set of all events H_1 in $\mathcal{F}_1$ for which 5.3 holds. By assumption, $\mathcal{D} \supset \mathcal{C}_1$ and $\Omega \in \mathcal{D}$, and the other two conditions for $\mathcal{D}$ to be a d-system on Ω are checked easily. It follows from the monotone class theorem that $\mathcal{D} \supset \sigma\mathcal{C}_1 = \mathcal{F}_1$. Repeating the procedure successively with $H_2, \ldots, H_n$ we see that 5.3 holds for all $H_1, \ldots, H_n$ in $\mathcal{F}_1, \ldots, \mathcal{F}_n$ respectively. In other words, 5.1 holds when the V_i are indicators. This is extended to arbitrary positive random variables V_i in $\mathcal{F}_i$ by using the form $V_i = \sum_{j=1}^{\infty} a_{ij} 1_{H_{ij}}$ (see Exercise I.2.27) and applying the monotone convergence theorem repeatedly. □

Independence of collections

The next proposition shows that independence survives groupings.

5.4 PROPOSITION. *Every partition of an independency is an independency.*

Proof. Let $\{\mathcal{F}_t : t \in T\}$ be an independency. Let $\{T_1, T_2, \ldots\}$ be a partition of T. Then, the subcollections $\mathcal{F}_{T_i} = \{\mathcal{F}_t : t \in T_i\}$, $i \in \mathbb{N}^*$, form a partition of the original independency. The claim is that they are independent, that is, $\{\mathcal{F}_{T_1}, \ldots, \mathcal{F}_{T_n}\}$ is an independency for each n. This follows from the preceding proposition: let $\mathcal{C}_i$ be a p-system of all events having the form of an intersection of finitely many events chosen from $\bigcup_{t \in T_i} \mathcal{F}_t$. Then, $\mathcal{C}_i$ generates $\mathcal{F}_{T_i}$ and $\Omega \in \mathcal{C}_i$, and 5.3 holds for the elements of $\mathcal{C}_1, \ldots, \mathcal{C}_n$ by the independence of the $\mathcal{F}_t$, $t \in T$. Thus, $\mathcal{F}_{T_1}, \ldots, \mathcal{F}_{T_n}$ are independent, and this is for arbitrary n. □

Pairwise independence

A collection of objects (like σ-algebras, random variables) are said to be *pairwise independent* if every pair of them is an independency. This is, of course, much weaker than being mutually independent. But independence can be checked by repeated checks for pairwise independence. We state this for a sequence of σ-algebras; it holds for a finite sequence as well, and therefore can be used to check the independency for arbitrary collections.

5.5 PROPOSITION. *The sub-σ-algebras $\mathcal{F}_1, \mathcal{F}_2, \ldots$ of $\mathcal{H}$ are independent if and only if $\mathcal{F}_{\{1,\ldots,n\}}$ and $\mathcal{F}_{n+1}$ are independent for all $n \geq 1$.*

Proof. Necessity is immediate from the last proposition. For sufficiency, suppose that $\mathcal{G}_n = \mathcal{F}_{\{1,\ldots,n\}} = \bigvee_{i=1}^n \mathcal{F}_i$ and $\mathcal{F}_{n+1}$ are independent for all n. Then, for $H_1, \ldots, H_m$ in $\mathcal{F}_1, \ldots, \mathcal{F}_m$ respectively, we can see that 5.3 holds by repeated applications of the independence of $\mathcal{G}_n$ and $\mathcal{F}_{n+1}$ for $n = m-1$, $m-2, \ldots, 1$ in that order. Thus, $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are independent by Proposition 5.2, and this is true for all $m \geq 2$. □

Independence of random variables

For each t in some index set T, let X_t be a random variable taking values in some measurable space $(E_t, \mathcal{E}_t)$. According to the general definitions above, the variables X_t are said to be *independent*, and the collection $\{X_t : t \in T\}$ is called an *independency*, if $\{\sigma X_t : t \in T\}$ is an independency.

Since a collection is an independency if and only if its every finite subset is an independency, we concentrate on the independence of a finite number of them, which amounts to taking $T = \{1, 2, \ldots, n\}$ for some integer $n \geq 2$.

5.6 PROPOSITION. *The random variables $X_1, \ldots, X_n$ are independent if and only if*

5.7
$$\mathbb{E}\, f_1 \circ X_1 \cdots f_n \circ X_n = \mathbb{E}\, f_1 \circ X_1 \cdots \mathbb{E}\, f_n \circ X_n$$

for all positive functions $f_1, \ldots, f_n$ in $\mathcal{E}_1, \ldots, \mathcal{E}_n$ respectively.

Proof. We need to show that 5.1 holds for all positive $V_1, \ldots, V_n$ in $\sigma X_1, \ldots, \sigma X_n$ respectively if and only if 5.7 holds for all positive $f_1, \ldots, f_n$ in $\mathcal{E}_1, \ldots, \mathcal{E}_n$ respectively. But this is immediate from Theorem 4.4: $V_i \in \sigma X_i$ if and only if $V_i = f_i \circ X_i$ for some f_i in $\mathcal{E}_i$. □

Let π be the joint distribution of $X_1, \ldots, X_n$, and let $\mu_1, \ldots, \mu_n$ be the corresponding marginals. Then, the left and the right sides of 5.7 are equal to, respectively,

$$\int_{E_1 \times \cdots \times E_n} \pi(dx_1, \ldots, dx_n)\, f_1(x_1) \cdots f_n(x_n)$$

and

$$\int_{E_1} \mu_1(dx_1)\, f_1(x_1) \int_{E_2} \cdots \int_{E_n} \mu_n(dx_n)\, f_n(x_n)$$

The equality of these two expressions for all positive $f_1, \ldots, f_n$ is equivalent to saying that $\pi = \mu_1 \times \cdots \times \mu_n$. We state this next.

5.8 Proposition. *The random variables $X_1, \ldots, X_n$ are independent if and only if their joint distribution is the product of their marginal distributions.*

Finally, a comment on functions of independent variables. In the language of Exercise 4.14, let $Y_1, \ldots, Y_n$ be determined by $X_1, \ldots, X_n$ respectively. Then $\sigma Y_i \subset \sigma X_i$ for $i = 1, \ldots, n$, and it follows from the definition of independency that $Y_1, \ldots, Y_n$ are independent if $X_1, \ldots, X_n$ are independent. We state this observation next.

5.9 Proposition. *Measurable functions of independent random variables are independent.*

Sums of independent random variables

Let X and Y be $\mathbb{R}^d$-valued independent random variables with distributions μ and ν respectively. Then, the distribution of (X, Y) is the product measure $\mu \times \nu$, and the distribution $\mu * \nu$ of $X + Y$ is given by

$$(\mu * \nu)f = \mathbb{E}f(X+Y) = \int_{\mathbb{R}} \mu(dx) \int_{\mathbb{R}} \nu(dy)\, f(x+y), \tag{5.10}$$

This distribution $\mu * \nu$ is called the *convolution* of μ and ν. See exercises below for more. Of course, since $X + Y = Y + X$, we have $\mu * \nu = \nu * \mu$. The convolution operation can be extended to any number of distributions.

Sums of random variables and the limiting behavior of such sums as the number of summands grows to infinity are of constant interest in probability theory. We shall return to such matters repeatedly in the chapters to follow. For the present, we describe two basic results, zero-one laws due to Kolmogorov and Hewitt-Savage.

Kolmogorov's 0-1 law

Let $(\mathcal{G}_n)$ be a sequence of sub-σ-fields of $\mathcal{H}$. We think of $\mathcal{G}_n$ as the information revealed by the nth trial of an experiment. Then, $\mathcal{T}_n = \bigvee_{m>n} \mathcal{G}_m$ is the information about the future after n, and $\mathcal{T} = \bigcap_n \mathcal{T}_n$ is that about the remote future. The last is called the *tail-σ-algebra*; it consists of events whose occurrences are unaffected by the happenings in finite time.

5.11 Example. Let $X_1, X_2, \ldots$ be real valued random variables, put $\mathcal{G}_n = \sigma X_n$ and $S_n = X_1 + \cdots + X_n$.

a) The event $\{\omega : \lim_n S_n(\omega) \text{ exists}\}$ belongs to $\mathcal{T}_n$ for every n and, hence, belongs to the tail-σ-algebra $\mathcal{T}$.

b) Similarly, $\{\limsup \frac{1}{n} S_n > b\}$ is unaffected by the first n variables, and this is true for all n, and hence this event belongs to $\mathcal{T}$.

c) But, $\{\limsup S_n > b\}$ is not in $\mathcal{T}$.

d) Let B be a Borel subset of $\mathbb{R}$. Let $\{X_n \in B \text{ i.o.}\}$, read X_n is in B *infinitely often*, be the set of ω for which $\sum_n 1_B \circ X_n(\omega) = +\infty$. This event belongs to $\mathcal{T}$.

e) The event $\{S_n \in B \text{ i.o.}\}$ is not in $\mathcal{T}$.

The following theorem, called Kolmogorov's 0-1 law, implies in particular that, if the X_n of the preceding example are independent, then each one of the events in $\mathcal{T}$ has probability equal to either 0 or 1.

5.12 THEOREM. *Let $\mathcal{G}_1, \mathcal{G}_2, \ldots$ be independent. Then, $\mathbb{P}(H)$ is either 0 or 1 for every event H in the tail $\mathcal{T}$.*

Proof. By Proposition 5.4 on partitions of independencies, $\{\mathcal{G}_1, \ldots, \mathcal{G}_n, \mathcal{T}_n\}$ is an independency for every n, which implies that so is $\{\mathcal{G}_1, \ldots, \mathcal{G}_n, \mathcal{T}\}$ for every n, since $\mathcal{T} \subset \mathcal{T}_n$. Thus, by definition, $\{\mathcal{T}, \mathcal{G}_1, \mathcal{G}_2, \ldots\}$ is an independency, and so is $\{\mathcal{T}, \mathcal{T}_0\}$ by Proposition 5.4 again. In other words, for H in $\mathcal{T}$ and $G \in \mathcal{T}_0$, we have $\mathbb{P}(H \cap G) = \mathbb{P}(H) \cdot \mathbb{P}(G)$, and this holds for $G = H$ as well because $\mathcal{T} \subset \mathcal{T}_0$. Thus, for H in $\mathcal{T}$, we have $\mathbb{P}(H) = \mathbb{P}(H) \cdot \mathbb{P}(H)$, which means that $\mathbb{P}(H)$ is either 0 or 1. □

As a corollary, assuming that the $\mathcal{G}_n$ are independent, for every random variable V in the tail-σ-algebra there is a constant c in $\bar{\mathbb{R}}$ such that $V = c$ almost surely. Going back to Example 5.11, for instance, $\limsup S_n/n$ is almost surely constant. In the same example, the next theorem will imply that the events $\{\limsup S_n > b\}$ and $\{S_n \in B \text{ i.o.}\}$ have probability 0 or 1, even though they are not in the tail $\mathcal{T}$, provided that we add to the independence of X_n the extra condition that they have the same distribution.

Hewitt-Savage 0-1 law

Let $X = (X_1, X_2, \ldots)$, where the X_n take values in some measurable space $(E, \mathcal{E})$. Let $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \ldots)$ be the filtration generated by X, that is, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for each n. Put $\mathcal{F}_\infty = \lim \mathcal{F}_n = \bigvee_n \mathcal{F}_n$, and recall from Theorem 4.4 and its sequel that $\mathcal{F}_\infty$ consists of random variables of the form $V = f \circ X$ with f in $\mathcal{E}^\infty$, and $\mathcal{F}_n$ consists of the variables of the form $V_n = f_n(X_1, \ldots, X_n) = \hat{f}_n \circ X$ with f_n in $\mathcal{E}^n$ (and appropriately defined $\hat{f}_n$).

By a *finite permutation* p is meant a bijection $p: \mathbb{N}^* \mapsto \mathbb{N}^*$ such that $p(n) = n$ for all but finitely many n. For such a permutation p, we write

5.13 $$X \circ p = (X_{p(1)}, X_{p(2)}, \ldots),$$

which is a re-arrangement of the entries of X. The notation extends to arbitrary random variables V in $\mathcal{F}_\infty$: if $V = f \circ X$ then $V \circ p = f \circ (X \circ p)$. It will

be useful to note that, if the X_n are independent and identically distributed, the probability laws of X and $X \circ p$ are the same, and hence, the distributions of V and $V \circ p$ are the same.

A random variable V in $\mathcal{F}_\infty$ is said to be *permutation invariant* if $V \circ p = V$ for every finite permutation p. An event in $\mathcal{F}_\infty$ is said to be permutation invariant if its indicator is such. These are variables like $V = \limsup S_n$ or events like $\{S_n \in B \text{ i.o.}\}$ in Example 5.11; they are unaffected by the re-arrangements of the entries of X by finite permutations.

The collection of all permutation invariant events is a σ-algebra which contains the tail-σ-algebra of X. The following, called Hewitt-Savage 0-1 law, shows that it is almost surely trivial (just as the tail) provided that the X_n are identically distributed in addition to being independent.

5.14 THEOREM. *Suppose that $X_1, X_2, \ldots$ are independent and identically distributed. Then, every permutation invariant event has probability 0 or 1. Also, for every permutation invariant random variable V there is a constant c in $\bar{\mathbb{R}}$ such that $V = c$ almost surely.*

Proof. It is sufficient to show that if $V : \Omega \mapsto [0,1]$ is a permutation invariant variable in $\mathcal{F}_\infty$, then $\mathbb{E}(V^2) = (\mathbb{E}V)^2$. Let V be such. By Proposition 4.9 there are V_n in $\mathcal{F}_n$, $n \geq 1$, such that each V_n takes values in $[0,1]$ and

5.15 $$\lim \mathbb{E}|V - V_n| = 0 \qquad \lim_n \mathbb{E}V_n = \mathbb{E}V,$$

the second limit being a consequence of the first.

Fix n. Let p be a finite permutation. The assumption about X implies that X and $X \circ p$ have the same probability law, which in turn implies that U and $U \circ p$ have the same distribution for every U in $\mathcal{F}_\infty$. Taking $U = V - V_n$, noting that $U \circ p = V \circ p - V_n \circ p = V - V_n \circ p$ by the invariance of V, we see that

5.16 $$\mathbb{E}|V - V_n \circ p| = \mathbb{E}|V - V_n|.$$

This is true, in particular, for the permutation $\hat{p}$ that maps $1, \ldots, n$ to $n+1, \ldots, 2n$ and vice-versa, leaving $\hat{p}(m) = m$ for $m > 2n$. We define $\hat{V}_n = V_n \circ \hat{p}$ and observe that, if $V_n = f_n(X_1, \ldots, X_n)$, then $\hat{V}_n = f_n(X_{n+1}, \ldots, X_{2n})$, which implies that V_n and $\hat{V}_n$ are independent and have the same distribution. Together with 5.16, this yields

5.17 $$\mathbb{E}V_n\hat{V}_n = (\mathbb{E}V_n)^2 , \qquad \mathbb{E}|V - \hat{V}_n| = \mathbb{E}|V - V_n|,$$

which in turn show that

5.18 $$|\mathbb{E}(V^2) - (\mathbb{E}V_n)^2| = |\mathbb{E}(V^2 - V_n\hat{V}_n)| \leq \mathbb{E}|V^2 - V_n\hat{V}_n| \leq 2\mathbb{E}|V - V_n|,$$

where the final step used (recalling $|V| \leq 1$ and $|V_n| \leq 1$)

$$|V^2 - V_n\hat{V}_n| = |(V - V_n)V + (V - \hat{V}_n)V_n| \leq |V - V_n| + |V - \hat{V}_n|,$$

and 5.17. Applying 5.15 to 5.18 yields the desired result that $\mathbb{E}V^2 = (\mathbb{E}V)^2$. □

5.19 EXAMPLE. *Random walks.* This is to provide a typical application of the preceding theorem. Returning to Example 5.11, assume further that $X_1, X_2, \ldots$ have the same distribution. Then, the stochastic process (S_n) is called a *random walk* on $\mathbb{R}$. The avoid the trivial case where $S_1 = S_2 = \ldots = 0$ almost surely, we assume that $\mathbb{P}\{X_1 = 0\} < 1$. Then, concerning the limiting behavior of the random walk, there are three possibilities, exactly one of which is almost sure:

i) $\lim S_n = +\infty$,

ii) $\lim S_n = -\infty$,

iii) $\liminf S_n = -\infty$, and $\limsup S_n = +\infty$.

Here is the argument for this. By the preceding theorem, there is a constant c in $\bar{\mathbb{R}}$ such that $\limsup S_n = c$ almost surely. Letting $\hat{S}_n = S_{n+1} - X_1$ yields another random walk $(\hat{S}_n)$ which has the same law as (S_n). Thus, $\limsup \hat{S}_n = c$ almost surely, which means that $c = c - X_1$. Since we excluded the trivial case when $\mathbb{P}\{X_1 = 0\} = 1$, it follows that c is either $+\infty$ or $-\infty$. Similarly, $\liminf S_n$ is either almost surely $-\infty$ or almost surely $+\infty$. Of the four combinations, discarding the impossible case when $\liminf S_n = +\infty$ and $\limsup S_n = -\infty$, we arrive at the result.

If the common distribution of the X_n is *symmetric*, that is, if X_1 and $-X_1$ have the same distribution (like the Gaussian with mean 0), then (S_n) and $(-S_n)$ have the same law, and it follows that the cases (i) and (ii) are improbable. So then, case (iii) holds almost surely.

Exercises

5.20 *Independence and functional independence.* Suppose that $(\Omega, \mathcal{H}, \mathbb{P}) = (\mathbb{B}, \mathcal{B}, \lambda) \times (\mathbb{B}, \mathcal{B}, \lambda)$, where $\mathbb{B} = [0, 1]$, $\mathcal{B} = \mathcal{B}(\mathbb{B})$ and λ is the Lebesgue measure on $\mathbb{B}$. For each $\omega = (\omega_1, \omega_2)$ in Ω, let $X(\omega) = f(\omega_1)$ and $Y(\omega) = g(\omega_2)$ for some Borel functions f and g on $\mathbb{B}$. Show that X and Y are independent.

5.21 *Independence and transforms.* Let X and Y be positive random variables. Then, X and Y are independent if and only if their joint Laplace transform is the product of their Laplace transforms, that is, if and only if

$$\mathbb{E}e^{-pX-qY} = \mathbb{E}e^{-pX}\,\mathbb{E}e^{-qY}\,, \qquad p, q \in \mathbb{R}_+.$$

Show this recalling that the joint Laplace transforms determine the joint distributions. A similar result holds for X and Y real-valued, but with characteristic functions. Obviously, these results can be extended to any finite number of variables.

5.22 *Sums of independent variables.* Let X and Y be independent real-valued random variables. Show that the characteristic function of $X + Y$ is the product of their characteristic functions. When X and Y are positive, the

same is true with Laplace transforms. When X and Y are positive integers, the same holds with generating functions. Use these to show the following.

a) If X has the Poisson distribution with mean a, and Y the Poisson distribution with mean b, then $X+Y$ has the Poisson distribution with mean $a+b$.

b) If X has the gamma distribution with shape index a and scale parameter c, and Y has the gamma distribution with shape index b and the same scale parameter c, then $X+Y$ has the gamma distribution with shape index $a+b$ and scale c.

c) If X has the Gaussian distribution with mean a and variance b and Y has the Gaussian distribution with mean c and variance d, then $X+Y$ has the Gaussian distribution with mean $a+c$ and variance $b+d$.

5.23 *Convolutions*

a) Let μ and ν be probability measures on $\mathbb{R}$, and let $\pi = \mu * \nu$ be defined by 5.10. Show that

$$\pi(B) = \int_{\mathbb{R}} \mu(dx)\, \nu(B-x)\,, \qquad B \in \mathcal{B}_{\mathbb{R}},$$

where $B - x = \{y - x :\ y \in B\}$.

b) Let λ be the Lebesgue measure on $\mathbb{R}$. Suppose that $\mu(dx) = \lambda(dx)\, p(x)$ and $\nu(dx) = \lambda(dx)\, q(x)$, $x \in \mathbb{R}$, for some positive Borel functions p and q. Show that, then, $\pi(dx) = \lambda(dx)\, r(x)$, where

$$r(x) = \int_{\mathbb{R}} dy\, p(y)\, q(x-y)\,, \qquad x \in \mathbb{R}.$$

Historically, then, r is said to be the convolution of the functions p and q, and the notation $r = p * q$ is used to indicate it.

c) Let μ and ν be as in the preceding case, but be carried by $\mathbb{R}_+$. Then, p and q vanish outside $\mathbb{R}_+$, and

$$r(x) = \int_0^x dy\, p(y)\, q(x-y)\,, \qquad x \in \mathbb{R}_+,$$

with $r(x) = 0$ for x outside $\mathbb{R}_+$.

Complements: Bernoulli sequences

5.24 *Bernoulli variables.* These are random variables that take the values 0 and 1 only. Each such variable is the indicator of an event, the event being named "success" to add distinction. Thus, if X is a Bernoulli variable, $p = \mathbb{P}\{X = 1\}$ is called the success probability, and then, $q = \mathbb{P}\{X = 0\} = 1 - p$ becomes the failure probability. Show that

$$\mathbb{E}\, X = \mathbb{E}\, X^2 = \cdots = p, \quad \operatorname{Var} X = pq, \quad \mathbb{E}\, z^X = q + pz.$$

5.25 *Bernoulli trials.* Let $X_1, X_2, \ldots$ be Bernoulli variables. It is usual to think of X_n as indicating the result of the nth trial in a sequence of trials: $X_n(\omega) = 1$ means that a "success" has occurred at the nth trial corresponding to the sequence described by the outcome ω. Often, it is convenient to assume that the trials occur at times $1, 2, 3, \ldots$. Then,

$$S_n = X_1 + \cdots + X_n$$

is the number of successes occuring during the time interval $[1, n]$. Assuming that $X_1, X_2, \ldots$ are independent and have the same success probability p (and the same failure probability $q = 1 - p$), show that

$$\mathbb{P}\{S_n = k\} = \frac{n!}{k!(n-k)!} p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

Hint: First compute $\mathbb{E}\ z^{S_n}$ using 5.24, and recall the binomial expansion $(a+b)^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} a^k b^{n-k}$. For this reason, the distribution of S_n is called the *binomial distribution.*

5.26 *Times of successes.* Let $X_1, X_2, \ldots$ be independent Bernoulli variables with the same success probability p. Define, for each k in $\mathbb{N}^*$, the time of kth success by

$$T_k(\omega) = \inf\{n \geq 1 : \ S_n(\omega) \geq k\}, \quad \omega \in \Omega.$$

Note that this yields $T_k(\omega) = +\infty$ if $S_n(\omega) < k$ for all n. Show that T_k is a random variable for each k in $\mathbb{N}^*$. Show that, for integers $n \geq k$,

$$\mathbb{P}\{T_k = n\} = \frac{(n-1)!}{(k-1)!(n-k)!} p^k q^{n-k}, \quad \mathbb{P}\{T_k \leq n\} = \sum_{j=k}^{n} \frac{n!}{j!(n-j)!} p^j q^{n-j}.$$

Show, in particular, that $T_k < \infty$ almost surely and, therefore, that $\lim S_n = +\infty$ almost surely.

5.27 *Waits between successes.* Let the X_n be as in 5.26. For $k \in \mathbb{N}^*$, define the waiting time $W_k(\omega)$ between the $(k-1)^{\text{th}}$ and k^{th} successes by letting $T_0(\omega) = 0$ and

$$W_k(\omega) = \begin{cases} T_k(\omega) - T_{k-1}(\omega) & \text{if } T_k(\omega) < \infty, \\ +\infty & \text{otherwise.} \end{cases}$$

For integers $i_1, \ldots, i_k$ in $\mathbb{N}^*$, express the event $\{W_1 = i_1, \ldots, W_k = i_k\}$ in terms of the variables X_k, compute the probability of the event in question, and conclude that $W_1, W_2, \ldots, W_k$ are independent random variables with the same distribution

$$\mathbb{P}\{W_k = i\} = pq^{i-1}, \quad i \in \mathbb{N}^*.$$

This distribution on $\mathbb{N}^*$ is called the *geometric distribution* with success probability p. Compute

$$\mathbb{E}\ W_k, \ \operatorname{Var} W_k, \ \mathbb{E}\ T_k, \ \operatorname{Var} T_k.$$

5.28 *Multinomial trials.* Let $X_1, X_2, \ldots$ be mutually independent random variables taking values in a finite set D, say $D = \{a, \ldots, d\}$, with

$$\mathbb{P}\{X_n = x\} = p(x), \quad x \in D.$$

For each x in D and ω in Ω, let $S_n(\omega, x)$ be the number of times that x appears in $(X_1(\omega), \ldots, X_n(\omega))$. Then, $S_n(x) : \omega \mapsto S_n(\omega, x)$ is a random variable for each n in $\mathbb{N}^*$ and each point x in D. Show that

$$\mathbb{P}\{S_n(a) = k(a), \ldots, S_n(d) = k(d)\} = \frac{n!}{k(a)! \cdots k(d)!} p(a)^{k(a)} \cdots p(d)^{k(d)}$$

for all $k(a), \ldots, k(d)$ in $\mathbb{N}$ with $k(a) + \cdots + k(d) = n$. This defines a probability measure on the simplex of all vectors $(k(a), \ldots, k(d))$ with $k(a) + \cdots + k(d) = n$; it is called a *multinomial distribution.*

5.29 *Empirical distributions.* Let $X_1, X_2, \ldots$ be mutually independent random variables taking values in some measurable space $(E, \mathcal{E})$and having the same distribution μ. Define

$$S_n(\omega, A) = \sum_{i=1}^{n} 1_A \circ X_i(\omega), \quad n \in \mathbb{N}, \omega \in \Omega, A \in \mathcal{E}.$$

Then, $A \mapsto S_n(\omega, A)$ is a counting measure on $(E, \mathcal{E})$whose atoms are the locations $X_1(\omega), \ldots, X_n(\omega)$, and $\frac{1}{n} S_n(\omega, A)$ defines a probability measure on $(E, \mathcal{E})$, called the *empirical distribution* corresponding to $X_1(\omega), \ldots, X_n(\omega)$. Writing $S_n(A)$ for the random variable $\omega \mapsto S_n(\omega, A)$, show that

$$\mathbb{P}\{S_n(A_1) = k_1, \ldots, S_n(A_m) = k_m\} = \frac{n!}{k_1! \cdots k_m!} \mu(A_1)^{k_1} \cdots \mu(A_m)^{k_m}$$

for every measurable partition $(A_1, \ldots, A_m)$ of E and integers $k_1, \ldots, k_m \geq 0$ summing to n.

5.30 *Inclusion-exclusion principle.* Let $X_1, \ldots, X_j$ be Bernoulli variables. Show that

$$\mathbb{P}\{X_1 = \cdots = X_j = 0\} = \mathbb{E} \sum_{Y} Y_1 \cdots Y_j$$

where the sum is over all j-tuples $Y = (Y_1, \ldots, Y_j)$ with each Y_i being either 1 or $-X_i$. Hint: The left side is the expectation of $(1 - X_1) \cdots (1 - X_j)$.

5.31 *Continuation.* For $X_1, \ldots, X_k$ Bernoulli, show that

$$\mathbb{P}\{X_1 = \cdots = X_j = 0, X_{j+1} = \cdots = X_k = 1\} = \mathbb{E} \sum_{Y} Y_1 \cdots Y_j X_{j+1} \cdots X_k$$

where the sum is over all Y as in 5.30.

5.32 *Probability law of a collection of Bernoullis.* Let I be an arbitrary index set. For each i in I, let X_i be a Bernoulli variable. Show that the probability law of $X = \{X_i : i \in I\}$ is specified by

$$\mathbb{E} \prod_{i \in J} X_i, \quad J \subset I,\ J \text{ finite} ;$$

in other words, knowing these expectations is enough to compute

$$\mathbb{P}\{X_i = b_i, i \in K\}$$

for every finite subset K of I and binary numbers b_i, $i \in K$.

5.33 *Independence of Bernoullis.* Show that X in 5.32 is an independency if and only if, for every finite $J \subset I$,

$$\mathbb{E} \prod_{i \in J} X_i = \prod_{i \in J} \mathbb{E}\, X_i.$$

Chapter III

Convergence

This chapter is devoted to various concepts of convergence: almost sure convergence, convergence in probability, convergence in L^p spaces, and convergence in distribution. In addition, the classical laws of large numbers and central limit theorems are presented in a streamlined manner.

1 Convergence of Real Sequences

The aim here is to review the concept of convergence in $\mathbb{R}$ and to bring together some useful results from analysis.

Let (x_n) be a sequence in $\mathbb{R}$, indexed by $\mathbb{N}^* = \{1, 2, \ldots\}$ as usual. Then,

$$\liminf x_n = \sup_m \inf_{n \geq m} x_n \ , \quad \limsup x_n = \inf_m \sup_{n \geq m} x_n$$

are well-defined numbers, possibly infinite. If these two numbers are equal to the same number x, then (x_n) is said to have the limit x, and we write $\lim x_n = x$ or $x_n \to x$ to indicate it. The sequence is said to be *convergent* in $\mathbb{R}$, or simply *convergent*, if the limit exists *and* is a real number.

Characterization

We start by introducing a notation for reasons of typographical convenience: for ε in $\mathbb{R}_+$, we let i_ε be the indicator of the interval (ε, ∞), that is,

1.1 $$i_\varepsilon(x) = 1_{(\varepsilon,\infty)}(x) = \begin{cases} 1 & \text{if } x > \varepsilon \ , \\ 0 & \text{if } x \leq \varepsilon \ . \end{cases}$$

Let (x_n) be a sequence in $\mathbb{R}$. It converges to x if and only if the sequence of positive numbers $|x_n - x|$ converges to 0. When x is known, it is simpler to work with the latter sequence because of its positivity.

Let (x_n) be a positive sequence. The classical statement of convergence would say that (x_n) converges to 0 if and only if for every $\varepsilon > 0$ there exists

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_3,

k such that $x_n \le \varepsilon$ for all $n \ge k$, in other words, for every $\varepsilon > 0$ the number of n for which $x_n > \varepsilon$ is finite. So,

1.2 $$x_n \to 0 \iff \sum_n i_\varepsilon(x_n) < \infty \text{ for every } \varepsilon > 0.$$

Since every term of the series on the right is either 0 or 1,

1.3 $$\sum_n i_\varepsilon(x_n) < \infty \iff \limsup_{n\to\infty} i_\varepsilon(x_n) = 0 \iff \lim_n i_\varepsilon(x_n) = 0.$$

Cauchy criterion

This is useful especially when there is no apriori candidate x for the limit. We omit the proof.

1.4 PROPOSITION. *The sequence (x_n) converges if and only if*

$$\lim_{m,n\to\infty} |x_m - x_n| = 0 \ ,$$

that is, for every $\varepsilon > 0$ there is k such that $|x_m - x_n| \le \varepsilon$ for all $m \ge k$ and $n \ge k$.

The following uses the Cauchy criterion together with some easy observations:

1.5 PROPOSITION. *If there exists a positive sequence (ε_n) such that*

$$\sum_n \varepsilon_n < \infty \ , \qquad \sum_n i_{\varepsilon_n}(|x_{n+1} - x_n|) < \infty \ ,$$

then (x_n) is convergent.

Proof. Let (ε_n) be such. Then, there is k such that $|x_{n+1} - x_n| \le \varepsilon_n$ for all $n \ge k$. Thus, for $n > m > k$,

$$|x_n - x_m| \le |x_n - x_{n-1}| + \cdots + |x_{m+1} - x_m| \le \varepsilon_m + \cdots + \varepsilon_{n-1} \le \sum_m^\infty \varepsilon_j \ .$$

By the assumed summability of (ε_n), the last member tends to 0 as $m \to \infty$. Hence, Cauchy criterion 1.4 applies, and (x_n) is convergent. □

Subsequences, selection principle

Let (x_n) be a sequence. A sequence (y_n) is said to be a *subsequence* of (x_n) if there exists an increasing sequence (k_n) in $\mathbb{N}$ with $\lim_n k_n = +\infty$ such that $y_n = x_{k_n}$ for each n. Regarding $\mathbb{N}$ as a sequence, we note that every such sequence (k_n) is a subsequence of $\mathbb{N}$. Denoting (k_n) by N, we shall

write $(x_n)_{n \in N}$ for the subsequence (y_n), and we shall say that (x_n) converges *along* N to x if the subsequence $(x_n)_{n \in N}$ converges to x.

Obviously, (x_n) tends to a limit x (infinite values are allowed for the limit) if and only if every subsequence of it has the same limit x. If (x_n) is bounded, it is always possible to extract a subsequence that is convergent; this is a fundamental property of the real number system. The following proposition, called the *selection principle*, is immediate from the observation that for every sequence (x_n) there is a subsequence whose limit is $\liminf x_n$ and a subsequence whose limit is $\limsup x_n$.

1.6 PROPOSITION. *If every subsequence that has a limit has the same value x for the limit, then the sequence tends to the same x (infinite values are allowed for x). If the sequence is bounded, and every convergent subsequence of it has the same limit x, then the sequence converges to x.*

Finally, we list the following lemma both as an illustration of using subsequences and as a useful little result that will be applicable a number of times.

1.7 LEMMA. *Let (x_n) be a sequence of positive numbers and put $\bar{x}_n = (x_1 + \cdots + x_n)/n$, the average of the first n entries. Let $N = (n_k)$ be a subsequence of $\mathbb{N}$ with* $\lim\ n_{k+1}/n_k = r > 0$. *If the sequence $(\bar{x}_n)$ converges along N to x, then*

$$x/r \leq \liminf \bar{x}_n \leq \limsup \bar{x}_n \leq r \cdot x \ .$$

Proof. For $n_k \leq n < n_{k+1}$, the positivity of the x_n yields

$$\frac{n_k}{n_{k+1}} \bar{x}_{n_k} \leq \bar{x}_n \leq \bar{x}_{n_{k+1}} \cdot \frac{n_{k+1}}{n_k} \ .$$

As $n \to \infty$, the integer k for which $n_k \leq n < n_{k+1}$ tends to $+\infty$, and our assumptions imply that the left-most member converges to x/r and the right-most to rx. □

Diagonal method

1.8 PROPOSITION. *Let $(x_{m,n})_{n \geq 1}$ be a bounded sequence for each integer $m \geq 1$. Then, there exists a subsequence N of $\mathbb{N}$ such that $(x_{m,n})_{n \in N}$ converges for every m.*

Proof. Since $(x_{1,n})$ is a bounded sequence, there is a subsequence N_1 of $\mathbb{N}$ such that $(x_{1,n})$ converges along N_1. The subsequence $(x_{2,n})_{n \in N_1}$ being bounded, N_1 has a subsequence N_2 such that $(x_{2,n})$ converges along N_2. Continuing in this manner, we obtain subsequences $N_1 \supset N_2 \supset \cdots$ such that, for each m, the sequence $(x_{m,n})$ converges along N_m.

Let n_m be the mth entry of N_m and define $N = (n_1, n_2, \ldots)$. Then, for each m, the tail $(n_m, n_{m+1}, \ldots)$ of N is a subsequence of N_m, and the convergence of $(x_{m,n})$ along N_m implies that $(x_{m,n})$ converges along N. □

If the sequences $N_1, N_2, \ldots$ in the proof were written one under the other, then N would be the diagonal of the resulting array. For this reason, the preceding proposition is called the diagonal method. It is useful in constructing limits of sequences of functions. The following is an application of some importance.

Helly's Theorem

Recall that a distribution function (in probabilistic terrain) is an increasing right-continuous mapping from $\mathbb{R}$ into $[0,1]$.

1.9 THEOREM. *For every sequence (c_n) of distribution functions there exists a subsequence (b_n) and a distribution function c such that* $\lim\ b_n(t) = c(t)$ *for every t at which c is continuous.*

Proof. We apply the diagonal method of the preceding proposition to $x_{m,n} = c_n(r_m)$, where (r_m) is an enumeration of the set of all rationals. Thus, $\mathbb{N}$ has a subsequence N such that, writing $(b_1, b_2, \ldots)$ for the subsequence $(c_n)_{n\in N}$, the limit $b(r) = \lim\ b_n(r)$ exists for every rational r. For each t in $\mathbb{R}$, define

$$c(t) = \inf\{b(r) : \quad r \text{ rational }, r > t\}.$$

It is clear that c is increasing. For each t in $\mathbb{R}$ and $\varepsilon > 0$, there is a rational $r > t$ such that $b(r) < c(t) + \varepsilon$, and we have $c(u) \le b(r)$ for all u in $[t, r)$. It follows that c is right-continuous as well. Thus c is a distribution function.

Let t be a point of continuity for c. Then, for every $\varepsilon > 0$ there is $s < t$ such that $c(s) > c(t) - \varepsilon$, and there is a rational $r > t$ such that $b(r) < c(t) + \varepsilon$. Pick a rational q in (s, t). We now have $s < q < t < r$ and $c(t) - \varepsilon < c(s) \le b(q) \le b(r) < c(t) + \varepsilon$. Since $b_n(q) \le b_n(t) \le b_n(r)$ for every n, and since $b_n(q) \to b(q)$ and $b_n(r) \to b(r)$, it follows that $\liminf b_n(t)$ and $\limsup b_n(t)$ are sandwiched between $c(t) - \varepsilon$ and $c(t) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, this shows that $\lim b_n(t) = c(t)$ as claimed. □

Kronecker's Lemma

This is a technical result which will be needed later. It relates convergence of averages to convergence of series.

1.10 LEMMA. *Let (x_n) be a sequence in $\mathbb{R}$. Let (a_n) be a strictly positive sequence increasing to $+\infty$. Put $y_n = \sum_1^n (x_k/a_k)$. If (y_n) is convergent, then*

$$\lim_n \frac{1}{a_n} \sum_1^n x_k = 0 \ .$$

Proof. Put $a_0 = y_0 = 0$. Note that $x_m = (y_m - y_{m-1})a_m$ and that

$$\sum_1^n x_m = \sum_{m=0}^{n-1} (a_{m+1} - a_m)(y_n - y_m) \ .$$

Suppose that $y_n \to y$. Fix $\varepsilon > 0$. Cauchy criterion 1.4 implies the existence of k such that $|y_n - y_m| \le \varepsilon$ for all $n, m \ge k$, and hence, on the right side, the partial sum over $m = k, \ldots, n-1$ becomes bounded in absolute value by $(a_n - a_k)\varepsilon \le a_n\varepsilon$. So,

$$|\frac{1}{a_n}\sum_{1}^{n} x_m| \le \varepsilon + \frac{1}{a_n}\sum_{m=0}^{k-1}(a_{m+1} - a_m)|y_n - y_m| \; .$$

On the right, the second term goes to 0 as $n \to \infty$, since $y_n \to y$ and $a_n \to \infty$; and $\varepsilon > 0$ can be taken to be arbitrarily small. □

2 Almost Sure Convergence

Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a probability space, and (X_n) is a sequence of real-valued random variables.

2.1 Definition. *The sequence* (X_n) *is said to be* almost surely convergent *if the numerical sequence* $(X_n(\omega))$ *is convergent for almost every* ω*; it is said to* converge to X *if* X *is an almost surely real-valued random variable and* $\lim X_n(\omega) = X(\omega)$ *for almost every* ω*.*

2.2 Remark. Since $\liminf X_n$ and $\limsup X_n$ are random variables, the set

$$\Omega_0 = \{\omega \in \Omega : \liminf X_n(\omega) = \limsup X_n(\omega) \in \mathbb{R}\}$$

is an event. The sequence (X_n) is almost surely convergent if and only if Ω_0 is almost sure, that is, $\mathbb{P}(\Omega_0) = 1$. This is the content of the definition above. Moreover, if Ω_0 is almost sure, then letting $X(\omega) = \lim X_n(\omega)$ for ω in Ω_0 and $X(\omega) = 0$ for $\omega \notin \Omega_0$, we obtain a real-valued random variable X such that $X_n \to X$ almost surely. Of course, if X' is another random variable such that $X = X'$ almost surely, then $X_n \to X'$ almost surely too.

Characterization theorem

2.3 Theorem. *The sequence* (X_n) *converges to* X *almost surely if and only if, for every* $\varepsilon > 0$,

2.4 $$\sum_n i_\varepsilon \circ |X_n - X| < \infty \quad \textit{almost surely.}$$

Proof. *Necessity.* Suppose $X_n \to X$ almost surely. Let Ω_0 be the almost sure set on which convergence holds, and let $Y_n = |X_n - X|$. Then, for each ω in Ω_0, by 1.2, $\sum_n i_\varepsilon \circ Y_n(\omega) < \infty$ for every $\varepsilon > 0$. Thus, for fixed $\varepsilon > 0$, 2.4 holds (and the almost sure set is Ω_0, which is further free of ε).

Sufficiency. Suppose that, for each $\varepsilon > 0$, the condition 2.4 holds. Let (ε_k) be a sequence strictly decreasing to 0. Let N_k be the random sum in 2.4

corresponding to ε_k. We have $\mathbb{P}\{N_k < \infty\} = 1$ by assumption for every k. Now, since $\varepsilon_{k+1} < \varepsilon_k$, we have $i_{\varepsilon_{k+1}} \geq i_{\varepsilon_k}$ and $N_{k+1} \geq N_k$. Thus, the events $\{N_k < \infty\}$ are shrinking to

$$\Omega_0 = \bigcap_k \{N_k < \infty\} = \{\omega \in \Omega : \sum_n i_\varepsilon \circ Y_n(\omega) < \infty \text{ for all } \varepsilon > 0\}.$$

By the sequential continuity for $\mathbb{P}$, we have $\mathbb{P}(\Omega_0) = \lim \mathbb{P}\{N_k < \infty\} = 1$, and for every ω in Ω_0 we have $X_n(\omega) \to X(\omega)$ in view of 1.2. □

Borel-Cantelli lemmas

The following three propositions provide sufficient conditions for almost sure convergence. They are referred to as Borel-Cantelli lemmas, because their main ingredient is the following classical result called Borel-Cantelli lemma.

2.5 Lemma. *Let (H_n) be a sequence of events. Then,*

$$\sum_n \mathbb{P}(H_n) < \infty \quad \Rightarrow \quad \sum_n 1_{H_n} < \infty \quad \textit{almost surely.}$$

Proof. Let N be the random variable whose finiteness is in question. By the monotone convergence theorem, $\mathbb{E}\, N = \sum_n \mathbb{P}(H_n)$. So, the claim is that if $\mathbb{E}\, N < \infty$ then $N < \infty$ almost surely, which is obvious. □

2.6 Proposition. *Suppose that*

$$\sum_n \mathbb{P}\{|X_n - X| > \varepsilon\} < \infty$$

for every $\varepsilon > 0$. Then, $X_n \to X$ almost surely.

Proof. The assumption implies, via the Borel-Cantelli lemma, that the condition 2.4 holds for every $\varepsilon > 0$. Thus, by Theorem 2.3, $X_n \to X$ almost surely. □

2.7 Proposition. *Suppose that there exists a sequence (ε_n) decreasing to 0 such that*

$$\sum_n \mathbb{P}\{|X_n - X| > \varepsilon_n\} < \infty\ .$$

Then, $X_n \to X$ almost surely.

Proof. By the Borel-Cantelli lemma, the assumption here implies that, for almost every ω, we have $|X_n(\omega) - X(\omega)| \leq \varepsilon_n$ for all but finitely many n, which in turn implies, since $\varepsilon_n \searrow 0$, that $X_n(\omega) \to X(\omega)$. □

2.8 PROPOSITION. *Suppose that there exists a sequence (ε_n) of strictly positive numbers such that*

$$\sum_n \varepsilon_n < \infty\ , \quad \sum_n \mathbb{P}\{|X_{n+1} - X_n| > \varepsilon_n\} < \infty\ .$$

Then, (X_n) converges almost surely.

Proof. By the Borel-Cantelli lemma, the assumption implies that the condition of Proposition 1.5 holds for the sequence $(x_n) = (X_n(\omega))$ for almost every ω. Thus, $(X_n(\omega))$ is convergent for almost every ω. □

Borel-Cantelli: divergence part

We interrupt the flow of this section to give a partial converse to Lemma 2.5. Let $B_1, B_2, \ldots$ be Bernoulli variables (that is, indicators of some events $H_1, H_2, \ldots$). If they are independent, then Kolmogorov's 0-1 law applies: either $\sum B_n < \infty$ almost surely, or $\sum B_n = \infty$; see Theorem II.5.12. Even without the independence assumption, Lemma 2.5 shows that the former case holds when $\sum \mathbb{E}B_n < \infty$. The following is a partial converse:

2.9 PROPOSITION. *Let $B_1, B_2, \ldots$ be Bernoulli variables.*

a) If $\sum \mathbb{E}B_n < \infty$, then $\sum B_n < \infty$ almost surely.

b) If $\sum \mathbb{E}B_n = +\infty$ and the B_n are pairwise independent, then $\sum B_n = +\infty$ almost surely.

Proof. The first claim is simply Lemma 2.5. To show (b), let $p_n = \mathbb{E}B_n$, $a_n = p_1 + \cdots + p_n$, $S_n = B_1 + \cdots + B_n$, $S = \lim S_n$. Assuming pairwise independence,

2.10 $$\mathrm{Var} S_n = \sum_1^n \mathrm{Var} B_i = \sum_1^n p_i(1 - p_i) \le \sum_1^n p_i = a_n\ .$$

Fix b in $(0, \infty)$. Since (a_n) increases to $+\infty$ by hypothesis, the numbers $a_n - \sqrt{ba_n}$ increase to $+\infty$. Thus, the event $\{S < \infty\}$ is the limit of the increasing sequence of events $\{S < a_n - \sqrt{ba_n}\}$, and (since $S_n \le S$) the latter event implies $\{S_n < a_n - \sqrt{ba_n}\}$, which in turn implies $\{|S_n - a_n| > \sqrt{ba_n}\}$. Hence,

$$\begin{aligned} \mathbb{P}\{S < \infty\} &= \lim \mathbb{P}\{S < a_n - \sqrt{ba_n}\} \\ &\le \limsup \mathbb{P}\{|S_n - a_n| > \sqrt{ba_n}\} \quad \le \quad \limsup (\mathrm{Var} S_n)/ba_n\ , \end{aligned}$$

the last inequality being Chebyshev's. In view of 2.10, this means that $\mathbb{P}\{S < \infty\} \le 1/b$, and letting $b \to \infty$ completes the proof. □

Cauchy criterion

As with its deterministic counterpart, this is useful where there is no apriori candidate for the limit of (X_n). In fact, the statement below is nothing but 1.4 stated for each ω in an almost sure event. No proof is needed.

2.11 THEOREM. *The sequence* (X_n) *is convergent almost surely if and only if* $\lim_{m,n\to\infty} |X_n - X_m| = 0$ *almost surely.*

To make the preceding practical we provide some details. For this purpose, let

2.12 $$Y_n = \sup_{i,j\geq n}|X_i - X_j| \ , \quad Z_n = \sup_k|X_{n+k} - X_n| \ .$$

The meaning of the Cauchy criterion is that $(X_n(\omega))$ is Cauchy if and only if $Y_n(\omega) \to 0$. And $Y_n(\omega) \to 0$ if and only if $Z_n(\omega) \to 0$, because $Z_n \leq Y_n \leq 2Z_n$. We put this as a lemma.

2.13 LEMMA. *The following are equivalent:* (X_n) *is almost surely convergent;* (Y_n) *converges to 0 almost surely;* (Z_n) *converges to 0 almost surely.*

2.14 PROPOSITION. *Suppose that*

$$\liminf_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left\{\sup_{k\leq m}|X_{n+k} - X_n| > \varepsilon\right\} = 0$$

for every $\varepsilon > 0$*. Then,* (X_n) *is convergent almost surely.*

Proof. Let $Z_{n,m}$ be the random variable that figures inside the event on the left. Note that $Z_{n,m}$ increases to Z_n as $m \to \infty$. Therefore, $i_\varepsilon \circ Z_{n,m} \to i_\varepsilon \circ Z_n$, which together with Fatou's lemma and bounded convergence theorem gives

$$\mathbb{E} \liminf i_\varepsilon \circ Z_n \leq \liminf \mathbb{E}\ i_\varepsilon \circ Z_n \leq \liminf_{n\to\infty} \lim_{m} \mathbb{E}\ i_\varepsilon \circ Z_{n,m} = 0 \ ,$$

the last equality being the hypothesis. Since a positive variable with 0 expectation is almost surely 0, we have shown that $\liminf i_\varepsilon \circ Z_n = 0$ almost surely. This is for every $\varepsilon > 0$. Since Y_n of 2.12 is bounded by $2Z_n$, it follows that, for every $\varepsilon > 0$, $\liminf i_\varepsilon \circ Y_n = 0$ almost surely. But (Y_n) is a decreasing sequence and $i_\varepsilon(y)$ is either 0 or 1. So, for every $\varepsilon > 0$, $\sum_n i_\varepsilon \circ Y_n < \infty$ almost surely, which implies via Theorem 2.4 that $Y_n \to 0$ almost surely. This in turn implies, through the preceding lemma, that (X_n) is almost surely convergent. □

Convergence in metric spaces

Let (E, d) be a metric space. Let $X_1, X_2, \ldots$ be random variables taking values in E. Then, (X_n) is said to converge to X almost surely provided that the real-valued variables $d(X_n, X)$ converge to 0 almost surely.

Complements

2.15 lim inf *and* lim sup *of events.* Let (H_n) be a sequence of events. Define

$$\liminf H_n = \bigcup_m \bigcap_{n \geq m} H_n \ , \quad \limsup H_n = \bigcap_m \bigcup_{n \geq m} H_n \ ;$$

these are again events. We have avoided using them.

a) Fix ω in Ω. Note that ω belongs to $\liminf H_n$ if and only if there exists m such that $\omega \in H_n$ for all $n \geq m$, and ω belongs to $\limsup H_n$ if and only if $\omega \in H_n$ for infinitely many n. For these reasons, some write $\{H_n \text{ ult.}\}$ for the lim inf and $\{H_n \text{ i.o.}\}$ for the lim sup, with "ult." an abreviation for "ultimately" and "i.o." for "infinitely often". These usages show a bizarre sense of notation, but the related usages $\{S_n \in B \text{ i.o.}\}$ for $\limsup\{S_n \in B\}$ and $\{X_n \in A \text{ ult.}\}$ for $\liminf\{X_n \in A\}$ are both proper and useful.

b) Show that

$$1_{\liminf H_n} = \liminf 1_{H_n} \ , \quad 1_{\limsup H_n} = \limsup 1_{H_n} \ .$$

c) Show that

$$\limsup H_n = \left\{ \sum_n 1_{H_n} = +\infty \right\} \ , \quad \liminf H_n = \left\{ \sum_n (1 - 1_{H_n}) < \infty \right\} \ .$$

3 Convergence in Probability

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let $X_1, X_2, \ldots, X$ be real-valued random variables.

3.1 Definition. *The sequence (X_n) is said to converge to X in probability if, for every $\varepsilon > 0$,*

$$\lim_n \mathbb{P}\{|X_n - X| > \varepsilon\} = 0 \ .$$

This mode of convergence is central to much of modern stochastics. In particular, stochastic calculus and stochastic differential equations employ convergence in probability as their basic mode of limit taking. This is so, because almost sure convergence is not as widely applicable as convergence in probability. Here is a concrete illustration.

3.2 Example. Let $\Omega = (0, 1]$, $\mathcal{H}$ the Borel σ-algebra on it, and $\mathbb{P}$ the Lebesgue measure. Let $X_1, X_2, X_3, X_4, X_5, X_6, \ldots$ be the indicators of $(0, 1]$, $(0, {}^1/_2]$, $({}^1/_2, 1]$, $(0, {}^1/_3]$, $({}^1/_3, {}^2/_3]$, $({}^2/_3, 1], \ldots$ respectively. Then, for arbitrary ε in $(0, 1)$, the probabilities $\mathbb{P}\{X_n > \varepsilon\}$ form the sequence $(1, {}^1/_2, {}^1/_2, {}^1/_3, {}^1/_3, {}^1/_3, \ldots)$, whose limit is obviously 0. So, $X_n \to 0$ in probability. But, for every ω in Ω, the sequence $(X_n(\omega))$ consists of zeros and ones

without end, which means that its limit inferior is 0, and limit superior 1. Thus, the set of ω for which $(X_n(\omega))$ is convergent is empty, that is, almost sure convergence fails miserably.

The following basic theorem characterizes convergence in probability in terms of almost sure convergence, and more.

3.3 THEOREM. *a) If (X_n) converges to X almost surely, then it converges to X in probability as well.*

b) If it converges to X in probability, then it has a subsequence that converges to X almost surely.

c) If its every subsequence has a further subsequence that converges to X almost surely, then it converges to X in probability.

Proof. To simplify the notation we assume that the (X_n) are positive and $X = 0$; this is nothing more than replacing $|X_n - X|$ with X_n. Recall that i_ε is the indicator of the interval (ε, ∞) and introduce

3.4 $$p_n = p_n(\varepsilon) = \mathbb{E}\, i_\varepsilon \circ X_n = \mathbb{P}\{X_n > \varepsilon\}\ .$$

a) Suppose that $X_n \to 0$ almost surely. Fix $\varepsilon > 0$. Then, $i_\varepsilon \circ X_n \to 0$ almost surely, which implies through the bounded convergence theorem that $p_n \to 0$. Thus, $X_n \to 0$ in probability.

b) Suppose that $X_n \to 0$ in probability. Let $\varepsilon_k = 1/k$, $k \geq 1$. Put $n_0 = 0$. For each $k \geq 1$, since $p_n(\varepsilon_k) \to 0$ as $n \to \infty$ by our assumption, there exists $n_k > n_{k-1}$ such that $p_n(\varepsilon_k) \leq 1/2^k$ for all $n \geq n_k$. Then,

$$\sum_k \mathbb{P}\{X_{n_k} > \varepsilon_k\} \leq \sum 1/2^k = 1\ .$$

Thus, Proposition 2.7 applies to the subsequence (X_{n_k}) to conclude that it converges to 0 almost surely.

c) Assume that every subsequence of (X_n) has a further subsequence that converges to 0 almost surely. Fix $\varepsilon > 0$ arbitrary. Consider the bounded sequence of numbers $p_n = p_n(\varepsilon)$. Let N be a subsequence of $\mathbb{N}$ along which (p_n) is convergent, let p be the limit. By assumption, N has a subsequence N' such that $X_n \to 0$ almost surely along N'. Then, by part (a) above, $p_n \to 0$ along N', which means that the limit p is 0. It follows from Proposition 1.6 that the original sequence (p_n) converges to 0, that is, $X_n \to 0$ in probability. □

Convergence and continuous functions

As an immediate application of the preceding theorem we list the following.

3.5 PROPOSITION. *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be continuous. If $X_n \to X$ in probability, then $f \circ X_n \to f \circ X$ in probability.*

Proof. Suppose that $X_n \to X$ in probability. Let N be a subsequence of $\mathbb{N}$. Since $X_n \to X$ in probability along N, Theorem 3.3b implies that N has a subsequence N' along which $X_n \to X$ almost surely, which in turn implies that $f \circ X_n \to f \circ X$ almost surely along N' by the continuity of f. Thus, Theorem 3.3c applies to the sequence $(f \circ X_n)$, and $f \circ X_n \to f \circ X$ in probability. □

Convergence and arithmetic operations

The method of the preceding proof shows that convergence in probability is preserved under arithmetical operations:

3.6 THEOREM. *Suppose that $X_n \to X$ and $Y_n \to Y$ in probability. Then, $X_n + Y_n \to X + Y$, and $X_n - Y_n \to X - Y$, and $X_n Y_n \to XY$, all in probability. Moreover, $X_n / Y_n \to X/Y$ in probability provided that, almost surely, Y and the Y_n are non-zero.*

Proof. We show the statement about the sum. Let N be a subsequence of $\mathbb{N}$. By Theorem 3.3b, N has a subsequence N' along which $X_n \to X$ almost surely. Since $Y_n \to Y$ in probability along N', there is a further subsequence N'' along which $Y_n \to Y$ almost surely. Of course, along N'', $X_n \to X$ almost surely. Hence, every subsequence N has a subsequence N'' along which $X_n + Y_n \to X + Y$ almost surely. Thus, by Theorem 3.3c, $X_n + Y_n \to X + Y$ in probability. □

Metric for convergence in probability

For real-valued random variables X and Y define

3.7 $$d(X, Y) = \mathbb{E}\left(|X - Y| \wedge 1\right) .$$

It is easy to check that $d(X, Y) = 0$ if and only if $X = Y$ almost surely and that $d(X, Y) + d(Y, Z) \geq d(X, Z)$. In other words, d is a metric on the space of all real-valued random variables provided that X and Y are identified as the same whenever $X = Y$ almost surely. The following shows that d is a metric for convergence in probability.

3.8 PROPOSITION. *The sequence (X_n) converges to X in probability if and only if $d(X_n, X) \to 0$ as $n \to \infty$.*

Proof. A simple drawing will show that, for ε in $(0, 1)$,

$$\varepsilon \; i_\varepsilon(x) \leq x \wedge 1 \leq \varepsilon + i_\varepsilon(x)$$

for all x in $\mathbb{R}_+$. Replacing x with $|X_n - X|$, taking expectations, and letting $n \to \infty$ completes the proof. □

Cauchy Criterion

3.9 THEOREM. *The sequence* (X_n) *converges in probability if and only if, for every* $\varepsilon > 0$,

3.10
$$\lim_{m,n\to\infty} \mathbb{P}\{|X_m - X_n| > \varepsilon\} = 0 \ .$$

Proof. a) Assume that (X_n) converges to some random variable X in probability. Pick $\varepsilon > 0$, let $\delta = \varepsilon/2$. Observe that

$$i_\varepsilon \circ |X_m - X_n| \le i_\delta \circ |X_m - X| + i_\delta \circ |X_n - X| \ ,$$

take expectations on both sides, and note that the terms on the right tend to 0 as $m, n \to \infty$ by our assumption that $X_n \to X$ in probability.

b) Assume that 3.10 holds for every $\varepsilon > 0$. Choose $\varepsilon_k = 1/2^k$, put $n_0 = 0$. For each $k \ge 1$, let $n_k > n_{k-1}$ be such that

$$\mathbb{P}\{|X_m - X_n| > \varepsilon_k\} \le 1/2^k$$

for all $m, n \ge n_k$. Put $Y_k = X_{n_k}$. Now the condition of Proposition 2.8 holds for (Y_k) and (ε_k), and, hence, (Y_k) converges almost surely; let X be its limit. Observe that, for $\varepsilon > 0$, with $\delta = \varepsilon/2$,

$$\mathbb{E}\ i_\varepsilon \circ |X_n - X| \le \mathbb{E}\ i_\delta \circ |X_n - X_{n_k}| + \mathbb{E}\ i_\delta \circ |Y_k - X| \ .$$

Now, as n and k tend to $+\infty$, the first term on the right side goes to 0 by the assumed 3.10, and the second term goes to 0 since $Y_k \to X$ almost surely and hence in probability. It follows that $X_n \to X$ in probability. □

Exercises

3.11 *Uniqueness of limits.* If $X_n \to X$ and $X_n \to Y$, both in probability, then $X = Y$ almost surely. Show.

3.12 *Effect of continuity.* Show that, if $X_n \to X$ in probability, then $\mathbb{E}\ f \circ X_n \to \mathbb{E}\ f \circ X$ for every bounded continuous function f on $\mathbb{R}$. Hint: Use the selection theorem 1.6 together with the proof of 3.5 on the bounded sequence $(\mathbb{E}\ f \circ X_n)$.

3.13 *Another characterization.* Show that $X_n \to X$ in probability if and only if, for some bounded continuous strictly increasing function f on $\mathbb{R}$, $f \circ X_n \to f \circ X$ in probability. Hint for sufficiency: Every such f is a homeomorphism of $\mathbb{R}$ onto some bounded interval (a, b), that is, f is a bijection from $\mathbb{R}$ onto (a, b), it is continuous, and its functional inverse is continuous.

3.14 *Continuation.* Show that $X_n \to X$ in probability if and only if $f \circ X_n \to f \circ X$ in probability for every bounded continuous function f on $\mathbb{R}$.

3.15 *Divergence.* The sequence (X_n) is said to *diverge* to $+\infty$ in probability if $\mathbb{P}\{X_n > b\} \to 1$ as $n \to \infty$ for every b in $\mathbb{R}_+$. If (X_n) diverges to $+\infty$ and (Y_n) converges to Y, both in probability, then $(X_n + Y_n)$ diverges to $+\infty$ in probability. Show.

3.16 *Convergence in metric spaces.* Let (E, r) be a metric space. Let X_1, $X_2, \ldots, X$ be E-valued random variables. Then, (X_n) is said to converge to X in probability if the real-valued sequence of random distances $r(X_n, X)$ converges to 0 in probability, that is, if

$$\lim_n \mathbb{P}\{r(X_n, X) > \varepsilon\} = 0$$

for every $\varepsilon > 0$. For $E = \mathbb{R}$, taking $r(x, y) = |x - y|$, we obtain Definition 3.1. Note that the metric defined by 3.7 is in fact $d(X, Y) = \mathbb{E}(r(X, Y) \wedge 1)$.

a) Show that Proposition 3.8 remains true with the present d for sequences in the metric space (E, r).

b) Show that Theorem 3.3 is true in metric spaces.

3.17 *Convergence in* $\mathbb{R}^d$. Fix the dimension $d \geq 1$. For x, y in $\mathbb{R}^d$, if $x = (x^1, \ldots, x^d)$ and $y = (y^1, \ldots, y^d)$, put

$$r(x, y) = \sum_{i=1}^{d} |x^i - y^i| \ .$$

Show that r is a metric on $\mathbb{R}^d$. Show that the sequence (X_n) in $\mathbb{R}^d$ converges in probability to X in $\mathbb{R}^d$ if and only if $X_n^i \to X^i$ in probability for each $i = 1, 2, \ldots, d$. Show that the same is true with the Euclidean distance

$$r(x, y) = \sqrt{\textstyle\sum_{i=1}^{d} (x^i - y^i)^2} \ .$$

4 Convergence in L^p

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. For p in $[1, \infty)$, the space L^p was introduced as the collection of all real-valued random variables X with $\mathbb{E}|X|^p < \infty$; see section II.3 for this and related concepts and notation.

4.1 Definition. *A sequence (X_n) is said to converge to X in L^p if every X_n is in L^p and X is in L^p and*

4.2 $$\lim_n \ \mathbb{E}\,|X_n - X|^p = 0 \ .$$

Recall the definition of L^p norm: $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. With this norm, L^p is a normed vector space once we identify as one random variable all random variables that are almost surely equal to each other (note that being almost surely equal is an equivalence relation). Convergence in L^p is the

ordinary concept of convergence in this normed vector space, because 4.2 is equivalent to

4.3 $$\lim_n \ \|X_n - X\|_p = 0 \ .$$

If (X_n) converges in L^p, then the limit X is unique up to equivalence: If Y is another random variable such that $X_n \to Y$ in L^p, then, by Minkowski's inequality,

$$\|X - Y\|_p \le \|X - X_n\|_p + \|X_n - Y\|_p \to 0 \ ,$$

and hence, $X = Y$ almost surely. Also, if the sequence converges to X in L^p, then it converges to the same X in probability: By Markov's inequality, for every $\varepsilon > 0$,

4.4 $$\mathbb{P}\{|X_n - X| > \varepsilon\} \le (\frac{1}{\varepsilon})^p \ \mathbb{E}\ |X_n - X|^p \to 0 \ .$$

The following example illustrates that there are not many converses to this. The full relationship between convergence in L^p and convergence in probability will be given in Theorems 4.6 and 4.9 below.

4.5 EXAMPLE. Consider the sequence (X_n) of Example 3.2. We had shown that it converges to 0 in probability. It converges to 0 in L^1 as well: the sequence of numbers $\mathbb{E}|X_n|$ is equal to $(1, {}^1/_2, {}^1/_2, {}^1/_3, {}^1/_3, {}^1/_3, {}^1/_4, \ldots)$, which converges to 0. A slight alteration produces a sequence $(\hat{X}_n)$ that converges in probability but not in L^1: Define $(\hat{X}_n) = (X_1, 2X_2, 2X_3, 3X_4, 3X_5, 3X_6, 4X_7, \ldots)$. Since $\mathbb{P}\{\hat{X}_n > \varepsilon\} = \mathbb{E}X_n$ for ε in $(0,1)$, we see that $\hat{X}_n \to 0$ in probability. But, now, $\mathbb{E}|\hat{X}_n| = 1$ for all n, and $(\hat{X}_n)$ does not converge to 0 in L^1; see 4.4 above that there could be no other limit in L^1. Finally, since $\hat{X}_n \to 0$ in probability, it has a subsequence (for instance $(\hat{X}_{n_k})$ with $n_k = 1 + (k^2-1)k^2/2$) that converges to 0 almost surely, but not in L^1.

Convergence, Cauchy, uniform integrability

The following is the main result of this section. We state it for $p = 1$ for reasons of simplicity. See Exercise 4.13 for arbitrary p.

4.6 THEOREM. *Let (X_n) be a sequence of real-valued random variables. For it, the following are equivalent:*

a) It converges in L^1.
b) It converges in probability and is uniformly integrable.
c) It is Cauchy for convergence in L^1, that is,

$$\lim_{m,n\to\infty} \mathbb{E}\ |X_m - X_n| = 0 \ .$$

Proof. We shall show that (a)⇒(c)⇒(b)⇒(a). i) Assume (a) and let X in L^1 be the limit. Then,

$$\mathbb{E}\,|X_m - X_n| \le \mathbb{E}\,|X_m - X| + \mathbb{E}\,|X - X_n| \to 0$$

as $m, n \to \infty$. Thus (c) holds.

ii) Assume (c). For every $\varepsilon > 0$, by Markov's inequality,

$$\mathbb{P}\{|X_m - X_n| > \varepsilon\} \le \frac{1}{\varepsilon}\,\mathbb{E}\,|X_m - X_n| \to 0$$

as $m, n \to \infty$. Thus, Theorem 3.9 applies, and the sequence converges in probability. To show that the sequence is uniformly integrable we use the ε-δ characterization of Theorem II.3.14: Fix $\varepsilon > 0$. Since the sequence is Cauchy in L^1, there exists an integer $k = k(\varepsilon)$ such that $\mathbb{E}\,|X_m - X_n| \le \varepsilon$ for all $m, n \ge k$. Thus, for every event H,

$$\mathbb{E}\,|X_n|1_H \le \mathbb{E}\,|X_n - X_k|1_H + \mathbb{E}\,|X_k|1_H \le \varepsilon + \mathbb{E}\,|X_k|1_H$$

for all $n \ge k$, and consequently,

$$\sup_n\ \mathbb{E}\,|X_n|1_H \le \varepsilon + \sup_{n\le k}\mathbb{E}\,|X_n|1_H\ .$$

On the right side, the finite collection $\{X_1, \ldots, X_k\}$ is uniformly integrable since the X_n are integrable; see Remark II.3.13. Hence, by Theorem II.3.14, there exists $\delta > 0$ such that $\mathbb{P}(H) \le \delta$ implies that the supremum over $n \le k$ is bounded by ε, and therefore supremum on the left side is bounded by 2ε. Finally, taking $H = \Omega$, we see that $\sup\ \mathbb{E}\,|X_n| < \infty$. Thus, the sequence is uniformly integrable and the implication (c)⇒(b) is proved.

iii) Assume (b). Let X be the limit. By Theorem 3.3 then, there is a subsequence (X_n') that converges to X almost surely, and Fatou's lemma yields

$$\mathbb{E}\,|X| = \mathbb{E}\ \liminf |X_n'| \le \liminf \mathbb{E}\,|X_n'| \le \sup_n \mathbb{E}\,|X_n|\ .$$

The supremum is finite by the assumed uniform integrability. Hence X is in L^1. To show that $X_n \to X$ in L^1, fix $\varepsilon > 0$, and let $H_n = \{|X_n - X| > \varepsilon\}$. Now, obviously,

$$\mathbb{E}\,|X_n - X| \le \varepsilon + \mathbb{E}\,|X_n - X|\ 1_{H_n}\ .$$

Since X is integrable and (X_n) is uniformly so, $(X_n - X)$ is uniformly integrable. Thus, there is $\delta > 0$ such that $\mathbb{P}(H_n) \le \delta$ implies that the expectation on the right is at most ε and $\mathbb{E}\,|X_n - X| \le 2\varepsilon$. Since $\mathbb{P}(H_n) \to 0$ by the assumed convergence in probability, this completes the proof that the sequence converges in L^1. □

Convergence of expectations, weak convergence in L^1

One reason for the popularity of convergence in L^1 is that it allows taking limits inside expectations: if $X_n \to X$ in L^1 then $\mathbb{E}\, X_n \to \mathbb{E}\, X$. This is a corollary to the following.

4.7 PROPOSITION. *If (X_n) converges to X in L^1, then*

4.8
$$\lim \mathbb{E}\, X_n Y = \mathbb{E}\, XY$$

for every bounded random variable Y.

Proof. Supposing that $|Y| \le b$, if $X_n \to X$ in L^1, then

$$|\mathbb{E}\, X_n Y - \mathbb{E}\, XY| \le \mathbb{E}\, |X_n Y - XY| \le b\, \mathbb{E}\, |X_n - X| \to 0\ . \qquad \square$$

A sequence (X_n) in L^1 is said to *converge weakly* in L^1 to X if 4.8 holds for every bounded variable Y. If so, then 4.8 holds for every Y that is almost surely bounded, that is, for every Y in L^∞. This mode of convergence introduces a new topology on L^1, often denoted by $\sigma(L^1, L^\infty)$.

A variation on the main results

4.9 THEOREM. *Suppose that (X_n) converges to X in probability. Then the following are equivalent for it:*

a) *It converges to X in L^1.*
b) *It is uniformly integrable.*
c) *It is a sequence in L^1, and $X \in L^1$, and $\mathbb{E}\, |X_n| \to \mathbb{E}\, |X|$.*

Proof. We have (a) $\iff$ (b) by Theorem 4.6, and (a) $\Rightarrow$ (c) since $|\,\mathbb{E}|X_n| - \mathbb{E}|X|\,| \le \mathbb{E}|X_n - X|$. To complete the proof we show that (c) $\Rightarrow$ (b). Assume (c), and note that the convergence of (X_n) to X in probability implies the convergence $|X_n| \to |X|$ in probability. So, there is no loss of generality in assuming further that the X_n and X are all positive.

Let $0 < a < b < \infty$. Define $f : \mathbb{R}_+ \mapsto [0, a]$ by setting $f(x)$ to be x on $[0, a]$, decrease continuously from a at a to 0 at b, and remain at 0 over (b, ∞). This f is bounded and continuous. Thus, by proposition 3.5, $f \circ X_n \to f \circ X$ in probability, and applying the implication (b) $\Rightarrow$ (c) here to the sequence $(f \circ X_n)$ we see that $\mathbb{E}\, f \circ X_n \to \mathbb{E}\, f \circ X$, and therefore, $\mathbb{E}\, (X_n - f \circ X_n) \to \mathbb{E}\, (X - f \circ X)$ since $\mathbb{E}\, X_n \to \mathbb{E}\, X$ by assumption. Recall that i_ε is the indicator of (ε, ∞), note that $x i_b(x) \le x - f(x)$ and $y - f(y) \le y i_a(y)$, replace x with X_n and y with X, and take expectations to conclude that

$$\limsup \mathbb{E}\, X_n\, i_b \circ X_n \le \mathbb{E}\, X\, i_a \circ X\ .$$

Fix $\varepsilon > 0$. By the integrability of X, the right side goes to 0 as $a \to \infty$; choose a so that the right side is less than $\varepsilon/2$. Then, definition of limit superior shows that there is m such that

4.10 $$\sup{}_{n>m}\mathbb{E}\ X_n\ i_b{\circ}X_n \le \varepsilon$$

for all $b \ge a+1$. Since $X_1, \ldots, X_m$ are in L^1 by assumption, by choosing a still larger if necessary, we ensure that 4.10 holds with the supremum taken over all n. Thus (X_n) is uniformly integrable. □

Exercises and complements

4.11 *Convergence of expectations.* Let $X_1, X_2, \ldots, X$ be in L^1. Show that $X_n \to X$ in L^1 if and only if $\mathbb{E}\ X_n 1_H \to \mathbb{E}\ X 1_H$ uniformly in H in $\mathcal{H}$, that is, if and only if

$$\lim_n \sup_{H\in\mathcal{H}} |\mathbb{E}\ X_n 1_H - \mathbb{E}\ X 1_H| = 0.$$

4.12 *Continuation.* If $X_n \to X$ in L^1, and $V_n \to V$ in L^1, and (V_n) is a bounded sequence, then $\mathbb{E}\ X_n V_n \to \mathbb{E}\ XV$. Show.

4.13 *Convergence in L^p, $p \in [1, \infty)$.* Show that the following are equivalent for every sequence (X_n):

a) The sequence converges in L^p.

b) The sequence is Cauchy in L^p, that is, $\mathbb{E}\ |X_m - X_n|^p \to 0$ as m, $n \to \infty$.

c) The sequence converges in probability and (X_n^p) is uniformly integrable.

Hint: Follow the proof of the basic theorem and use the generalization $|x+y|^p \le 2^{p-1}(|x|^p + |y|^p)$ of the triangle inequality.

4.14 *Weak convergence in L^1.* A sequence is uniformly integrable if and only if its every subsequence has a further subsequence that converges weakly in L^1. This is a deep result.

5 Weak Convergence

This section is about the convergence of sequences of probability measures on a given topological space. We limit ourselves to the space $\mathbb{R}$ and make a few remarks for the case of general spaces.

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let $\mu_1, \mu_2, \ldots, \mu$ be probability measures on $\mathbb{R}$. Let $X_1, X_2, \ldots, X$ be $\mathbb{R}$-valued random variables whose respective distributions are $\mu_1, \mu_2, \ldots, \mu$. Finally, let $\mathbb{C}_b = C_b(\mathbb{R}, \mathbb{R})$, the collection of all bounded continuous functions from $\mathbb{R}$ into $\mathbb{R}$.

5.1 Definition. *The sequence* (μ_n) *is said to* converge weakly *to* μ *if* $\lim \mu_n f = \mu f$ *for every* f *in* $\mathbb{C}_b$. *The sequence* (X_n) *is said to* converge in distribution *to* X *if* (μ_n) *converges to* μ *weakly, that is, if*

$$\lim \quad \mathbb{E}\ f\circ X_n = \mathbb{E}\ f\circ X$$

for every f *in* $\mathbb{C}_b$.

5.2 Remarks. a) Convergence in probability (or in L^1, or almost surely) implies convergence in distribution. To see this, let $X_n \to X$ in probability; and let $f \in \mathbb{C}_b$. Then, by Theorem 3.3, every subsequence of $\mathbb{N}$ has a further subsequence N such that $X_n \to X$ along N almost surely, $f\circ X_n \to f\circ X$ along N almost surely by the continuity of f, and $\mathbb{E}\ f\circ X_n \to \mathbb{E}\ f\circ X$ along N by the bounded convergence theorem. By the selection principle 1.6, then, $\mathbb{E}\ f\circ X_n \to \mathbb{E}\ f\circ X$.

b) There is a partial converse: Suppose that (X_n) converges to X in distribution *and* $X = x_0$ for some fixed point x_0. Then, in particular, for f defined by letting $f(x) = |x - x_0| \wedge 1$,

$$\mathbb{E}\ |X_n - X| \wedge 1 = \mathbb{E}\ f\circ X_n \to \mathbb{E}\ f\circ X = f(x_0) = 0\ .$$

Thus, $X_n \to X = x_0$ in probability by Proposition 3.8.

c) In general, convergence in distribution implies no other kind. For example, if the X_n are independent and have the same distribution as X, then $\mu_1 = \mu_2 = \cdots = \mu$ and (X_n) converges in distribution to X. But it does not converge in probability except in the trivial case where $X_1 = X_2 = \cdots = X = x_0$ almost surely for some fixed point x_0.

d) As the preceding remark illustrates, convergence of (X_n) in distribution has little to do with the convergence of (X_n) as a sequence of functions. Convergence in distribution is merely a convenient turn of phrase for the weak convergence of the corresponding probability measures.

5.3 Examples. a) *Convergence to Lebesgue measure.* Let μ_n be the probability measure that puts mass $1/n$ at each of the points $1/n, 2/n, \ldots, n/n$. Then, for f in $\mathbb{C}_b$,

$$\mu_n f = \sum_{k=1}^{n} \frac{1}{n} f(\frac{k}{n}) \to \int_0^1 du\ f(u) = \lambda f\ ,$$

where λ denotes the Lebesgue measure on $[0,1]$. Thus, (μ_n) converges weakly to λ.

b) *Quantile functions.* Let $q : (0,1) \mapsto \mathbb{R}$ be the quantile function corresponding to μ and define q_n similarly for μ_n; see Exercise II.1.18 *et seq.* Then, $\mu = \lambda\circ q^{-1}$ and $\mu_n = \lambda\circ q_n^{-1}$, where λ is the Lebesgue measure on $(0,1)$. Suppose that $q_n(u) \to q(u)$ for λ-almost every u in $(0,1)$. For f in $\mathbb{C}_b$, then, $f\circ q_n(u) \to f\circ q(u)$ for λ-almost every u in $(0,1)$ and therefore

$$\mu_n f = \lambda(f \circ q_n) \to \lambda(f \circ q) = \mu f$$

by the bounded convergence theorem. Hence, if (q_n) converges to q almost everywhere on $(0, 1)$, then (μ_n) converges to μ weakly. We shall see later in Proposition 5.7 that the converse holds as well.

Characterization theorem

The following basic theorem characterizes weak convergence of (μ_n) to μ in terms of the convergence of numbers $\mu_n(A)$. Here, ∂A denotes the boundary of A, $\bar{A}$ its closure, $\mathring{A}$ its interior (so that $\bar{A} = A \cup \partial A$, $\mathring{A} = A \setminus \partial A$, $\bar{A} \setminus \mathring{A} = \partial A$). We also write $d(x, y)$ for the usual distance, $|x - y|$, between the points of $\mathbb{R}$.

5.4 THEOREM. *The following are equivalent:*

a) (μ_n) *converges weakly to* μ.
b) $\limsup \mu_n(A) \le \mu(A)$ *for every closed set* A.
c) $\liminf \mu_n(A) \ge \mu(A)$ *for every open set* A.
d) $\lim \mu_n(A) = \mu(A)$ *for every Borel set* A *with* $\mu(\partial A) = 0$.

Proof. We shall show that (a) $\Rightarrow$ (b) $\iff$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a). i) Assume (a). Let A be closed. Let $d(x, A) = \inf\{d(x, y) : y \in A\}$ and let $A_\varepsilon = \{x : d(x, A) < \varepsilon\}$. Since A is closed, A_ε shrinks to A and, hence, $\mu(A_\varepsilon) \searrow \mu(A)$ as $\varepsilon \searrow 0$. Thus, to show (b), it is sufficient to show that

$$\limsup \mu_n(A) \le \mu(A_\varepsilon) \tag{5.5}$$

for every $\varepsilon > 0$. To this end, fix $\varepsilon > 0$ and define $f(x) = (1 - d(x, A)/\varepsilon) \vee 0$. Then, f is continuous and bounded, and $1_A \le f$, and $f \le 1_{A_\varepsilon}$. Hence, $\mu_n(A) \le \mu_n f$, $\mu f \le \mu(A_\varepsilon)$, and $\mu_n(f) \to \mu f$, which show that 5.5 holds.

ii) We have (b) $\iff$ (c), because the complements of open sets are closed and vice versa, and $\liminf(1 - r_n) = 1 - \limsup r_n$ for every sequence (r_n) in $[0, 1]$.

iii) Suppose that (c) and therefore (b) hold. Let A be a Borel set. Since $\bar{A} \supset A \supset \mathring{A}$, using (b) and (c), we obtain

$$\mu(\bar{A}) \ge \limsup \mu_n(\bar{A}) \ge \limsup \mu_n(A) \ge \liminf \mu_n(A) \ge \liminf \mu_n(\mathring{A}) \ge \mu(\mathring{A}).$$

If $\mu(\partial A) = 0$, then $\mu(\bar{A}) = \mu(\mathring{A})$, and all the inequalities here become equalities and show that $\lim \mu_n(A) = \mu(A)$, So, (d) holds.

iv) Suppose that (d) holds. Let $f \in \mathbb{C}_b$. Choose a and b in $\mathbb{R}$ such that $a < f < b$. Fix $\varepsilon > 0$ arbitrary. Considering the probability measure $\mu \circ f^{-1}$ on (a, b), pick $a = a_0 < a_1 < \cdots < a_k = b$ such that $a_i - a_{i-1} \le \varepsilon$ for all i and no a_i is an atom for $\mu \circ f^{-1}$ (this is possible since a probability measure has at most countably many atoms). Let $A_i = f^{-1}(a_{i-1}, a_i]$, define

$$g = \sum_1^k a_{i-1} 1_{A_i} \ , \quad h = \sum_1^k a_i 1_{A_i}$$

and observe that

5.6 $$f - \varepsilon \le g \le f \le h \le f + \varepsilon.$$

If $x \in \partial A_i$ then $f(x)$ is either a_{i-1} or a_i, neither of which is an atom for $\mu \circ f^{-1}$. Thus, $\mu(\partial A_i) = 0$ and it follows from assuming (d) that $\mu_n(A_i) \to \mu(A_i)$ as $n \to \infty$ for $i = 1, \ldots, k$. Thus, $\mu_n g \to \mu g$ and $\mu_n h \to \mu h$, and 5.6 yields

$$\begin{aligned} \mu f - \varepsilon \quad \le \quad \mu g = \lim \mu_n g \quad &\le \quad \liminf \mu_n f \\ &\le \quad \limsup \mu_n f \quad \le \quad \lim \mu_n h = \mu h \quad \le \quad \mu f + \varepsilon. \end{aligned}$$

In other words, limit inferior and limit superior of the sequence $(\mu_n f)$ are sandwiched between the numbers $\mu f - \varepsilon$ and $\mu f + \varepsilon$ for arbitrary $\varepsilon > 0$. So, $\mu_n f \to \mu f$ as needed to show that (a) holds. □

Uniqueness of weak limits and equality of measures

Let μ and ν be probability measures on $\mathbb{R}$ and suppose that $\mu f = \nu f$ for every f in $\mathbb{C}_b$. Then, as was pointed in Exercise II.2.34, $\mu = \nu$. Here is a simple proof: Let $\mu_n = \nu$ for all n; then (μ_n) converges weakly to μ; and it follows from the preceding theorem that $\nu(A) \ge \mu(A)$ for every open set A. Reversing the roles of μ and ν, we conclude that $\mu(A) = \nu(A)$ for every open set A. Since the open sets form a p-system that generates the Borel σ-algebra, it follows from Proposition I.3.7 that $\mu = \nu$.

Consequently, if (μ_n) converges weakly to μ and also to ν, then $\mu f = \nu f$ for every f in $\mathbb{C}_b$, and hence $\mu = \nu$. Therefore, (μ_n) has at most one weak limit.

Convergence of quantiles and distribution functions

Let $\mu_1, \mu_2, \ldots, \mu$ be probability measures on $\mathbb{R}$ as before. Let $c : \mathbb{R} \mapsto [0,1]$ be the distribution function corresponding to μ, and $q : (0,1) \mapsto \mathbb{R}$ the corresponding quantile function, and let c_n and q_n be associated with μ_n similarly, see Exercise II.2.18 *et seq.* for the definitions and various connections.

5.7 PROPOSITION. *The following are equivalent:*

a) (μ_n) *converges to* μ *weakly.*
b) $c_n(x) \to c(x)$ *for every continuity point* x *of* c.
c) $q_n(u) \to q(u)$ *for every continuity point* u *of* q.

Proof. Suppose that (a) holds. Let x be a point of continuity for c. Then, $\mu\{x\} = c(x) - c(x-) = 0$. Since the boundary of $(-\infty, x]$ is $\{x\}$, it follows from the characterization theorem above that $c_n(x) = \mu_n(-\infty, x] \to \mu(-\infty, x] = c(x)$. Thus, (a) $\Rightarrow$ (b).

Suppose that (b) holds. Let u be a point of continuity for q. Set $x = q(u)$, fix $\varepsilon > 0$, pick y in $(x - \varepsilon, x)$ and z in $(x, x + \varepsilon)$ to be continuity points for c.

Since q is continuous at u, the function c does not remain flat at level u, and hence $c(y) < u < c(z)$. Since $c_n(y) \to c(y)$ by assumption, we have $c_n(y) < u$ and thus $q_n(u) > y > x - \varepsilon$ for all but finitely many n, which means that $\liminf q_n(u) > x - \varepsilon$. Similarly, $c_n(z) > u$ and therefore $q_n(u) \le z < x + \varepsilon$ for all but finitely many n, which means that $\limsup q_n(u) < x + \varepsilon$. Thus, $\lim q_n(u) = x = q(u)$; and (b) $\Rightarrow$ (c).

Suppose that (c) holds. Since q is increasing right-continuous, it has at most countably many discontinuities (all jumps), and thus q is continuous Lebesgue-almost everywhere on $(0, 1)$. It follows as in Example 5.3b that (μ_n) converges weakly to μ, that is, (c) $\Rightarrow$ (a). □

Almost sure representations of weak convergence

The equivalence of (a) and (c) in the preceding theorem can be exploited further. The basic point is the relationship between quantiles and distributions: namely, as shown in II.2.20, μ is the image of the Lebesgue measure on $(0, 1)$ under the mapping $q : (0, 1) \mapsto \mathbb{R}$, and similarly for μ_n and q_n.

5.8 Theorem. *The sequence (μ_n) converges weakly to μ if and only if there exist random variables $Y_1, Y_2, \ldots, Y$ on some probability space $(\Omega', \mathcal{H}', \mathbb{P}')$ such that the distribution of Y_n is μ_n for each n, the distribution of Y is μ, and (Y_n) converges to Y almost surely on $(\Omega', \mathcal{H}', \mathbb{P}')$.*

Proof. Sufficiency. If such random variables exist, then by Remark 5.2a, (Y_n) converges to Y in distribution, which is equivalent to saying that (μ_n) converges weakly to μ.

Necessity. Suppose that (μ_n) converges weakly to μ. Then, the preceding Proposition 5.7 implies that the corresponding quantile functions q_n converge to q Lebesgue-almost everywhere on $(0, 1)$. Let $\Omega' = (0, 1)$, $\mathcal{H}'$ the Borel σ-algebra on it, and $\mathbb{P}'$ the Lebesgue measure. Define $Y_n = q_n$ and $Y = q$ on Ω'. Then , the distribution of Y_n is $\mathbb{P}' \circ q_n^{-1} = \mu_n$ and the distribution of Y is $\mathbb{P}' \circ q^{-1} = \mu$, and $Y_n(w) \to Y(w)$ for $\mathbb{P}'$-almost every w in Ω', (see also Example 5.3b). □

The following translates the preceding theorem by using the euphemism "convergence in distribution" instead of "weak convergence".

5.9 Corollary. *The sequence (X_n) converges in distribution to X if and only if there exist random variables $Y_1, Y_2, \ldots, Y$ (on some probability space) such that Y_n has the same distribution as X_n for each n, and Y has the same distribution as X, and (Y_n) converges almost surely to Y.*

This corollary and the preceding theorem are called *Skorokhod representations* especially when the space $\mathbb{R}$ here is replaced with some metric space, in which case the construction of the "quantile functions" is not as easy. However, the basic idea remains the same as with representation of measures described in Theorem I.5.4 and Exercise I.5.15; See Exercise 5.29 for more on this.

Such representations elevate convergence in distribution to the level of almost sure convergence in situations where the desired results concern only the distributions μ_n and μ. Here is an illustration of the method.

5.10 Proposition. *Suppose that* (X_n) *converges in distribution to* X. *Then the following are equivalent:*

a) (X_n) *is uniformly integrable.*
b) *The* X_n *and* X *are integrable and* $\mathbb{E}\,|X_n| \to \mathbb{E}\,|X|$.

Proof. Let (Y_n) and Y be as in the last corollary. Since (a) and (b) are in fact statements about the marginal distributions μ_n and μ, it is sufficient to show that (a) and (b) remain equivalent when the X_n and X are replaced with the Y_n and Y. But, then, the equivalence is immediate from Theorem 4.9. □

It is worth noting the absence here of the third statement in 4.9, the one about the convergence of (X_n) to X in L^1. This is because convergence in L^1 concerns the sequence of joint distributions π_n of the pairs (X_n, X), and we have no guarantee that the joint distribution of Y_n and Y is π_n for each n.

Convergence of image measures

Let $h : \mathbb{R} \mapsto \mathbb{R}$ be Borel, and suppose that the set D_h of its discontinuity points is a Borel set.

5.11 Proposition. *If* (μ_n) *converges weakly to* μ *and if* $\mu(D_h) = 0$, *then* $(\mu_n \circ h^{-1})$ *converges weakly to* $\mu \circ h^{-1}$. *If* (X_n) *converges in distribution to* X *and* $\mathbb{P}\{X \in D_h\} = 0$, *then* $(h \circ X_n)$ *converges in distribution to* $h \circ X$.

Proof. Let the Y_n and Y be as in the representation theorem 5.8. The assumption that $\mu(D_h) = 0$ implies that, almost surely, Y takes values outside D_h. Hence, almost sure convergence of (Y_n) to Y implies the almost sure convergence of $(h \circ Y_n)$ to $h \circ Y$. By Theorem 5.8, this means that $(\mu_n \circ h^{-1})$ converges weakly to $\mu \circ h^{-1}$. This proves the first statement; the second is a translation of the first. □

Tightness and Prohorov's Theorem

For each f in $\mathbb{C}_b$, the sequence $(\mu_n f)$ is bounded. In view of Proposition 1.6, if every subsequence of it has a further subsequence that converges to μf, then $\mu_n f \to \mu f$. It follows that, if every subsequence of (μ_n) has a further subsequence that converges weakly to μ, then (μ_n) converges weakly to μ. The next theorem, called Prohorov's, goes a long way toward making this idea work. But, first, a new concept:

5.12 Definition. *The sequence* (μ_n) *is said to be* tight *if for every* $\varepsilon > 0$ *there is a compact set* K *such that* $\mu_n(K) > 1 - \varepsilon$ *for all* n.

5.13 THEOREM. *If (μ_n) is tight then every subsequence of it has a further subsequence that is weakly convergent.*

Proof. Let (μ_n) be tight. Let (c_n) be the corresponding sequence of distribution functions. Let N be a subsequence of $\mathbb{N}$. By Helly's theorem 1.7, there is a distribution function c and a subsequence N' such that (c_n) converges along N' to c pointwise on the continuity set of c. To the distribution function c, there corresponds a measure μ, which we shall show presently to be a probability measure. Then, Proposition 5.7 implies that (μ_n) converges weakly to μ, which concludes the proof.

To show that the measure μ is a probability measure on $\mathbb{R}$, we need to show that $c(-\infty) = 0$ and $c(+\infty) = 1$. To this end, fix $\varepsilon > 0$. Since (μ_n) is tight, there is a closed interval $[a, b]$ such that $\mu_n[a, b] > 1-\varepsilon$ for all n. Then, $c_n(x) \le \varepsilon$ for all $x < a$, and $c_n(y) > 1 - \varepsilon$ for all $y > b$, these being true for all n and therefore for all n in N'. It follows that $c(x) \le \varepsilon$ for $x < a$ and $c(y) > 1 - \varepsilon$ for $y > b$. This implies that $c(-\infty) \le \varepsilon$ and $c(+\infty) > 1 - \varepsilon$ for all $\varepsilon > 0$ and hence the desired end. □

Convergence of Fourier transforms

Let (μ_n) be as before, and let (f_n) be the corresponding sequence of Fourier transforms, that is, for each n,

5.14 $$f_n(r) = \int_{\mathbb{R}} \mu_n(dx)e^{irx} , \qquad r \in \mathbb{R} .$$

The next theorem connects the convergence of (μ_n) to that of (f_n):

5.15 THEOREM. *The sequence (μ_n) is weakly convergent if and only if*

5.16 $$\lim_n f_n(r) = f(r)$$

exists for every r in $\mathbb{R}$ and the function f is continuous at 0. Moreover, then, f is the Fourier transform of a probability measure μ on $\mathbb{R}$, and μ is the weak limit of (μ_n).

Proof. *Necessity.* Assume that (μ_n) converges weakly to a probability measure μ on $\mathbb{R}$. Then, since $g : x \mapsto \cos(rx)$ and $h : x \mapsto \sin(rx)$ are in $\mathbb{C}_b$, $\mu_n g \to \mu g$ and $\mu_n h \to \mu h$ and hence $f_n(r) = \mu_n g + i\mu_n h \to \mu g + i\mu h = f(r)$, where f is the Fourier transform of μ. Then, f is continuous at 0 since it is the transform of a probability measure on $\mathbb{R}$.

Sufficiency. Suppose that the limits in 5.16 exist and f is continuous at 0. By the bounded convergence theorem,

$$\lim_n \frac{1}{b}\int_{-b}^{b} dr|1 - f_n(r)| = \frac{1}{b}\int_{-b}^{b} dr|1 - f(r)| ,$$

and the right side goes to 0 as $b \to 0$, because $f(r) \to f(0) = \lim f_n(0) = 1$ as $r \to 0$ by the assumed continuity of f at 0. Hence, for every $\varepsilon > 0$ there is $b > 0$ such that the right side is less than $\varepsilon/2$, and thus

5.17 $$\frac{1}{b}\int_{-b}^{b} dr|1 - f_n(r)| \le \varepsilon$$

for all but finitely many n.

On the other hand, using Fubini's Theorem and 5.14,

5.18 $$\begin{aligned} \frac{1}{b}\int_{-b}^{b} dr(1 - f_n(r)) &= \int_{\mathbb{R}} \mu_n(dx)\frac{1}{b}\int_{-b}^{b} dr\ (1 - e^{irx}) \\ &= \int_{\mathbb{R}} \mu_n(dx)\ 2(1 - \frac{\sin bx}{bx}) \\ &\ge \int_{\mathbb{R}} \mu_n(dx)\ 1_{(2,\infty)}(b|x|) = 1 - \mu_n[-\frac{2}{b}, \frac{2}{b}]\ , \end{aligned}$$

where the inequality is justified by noting that $1 - (\sin y)/y$ is always positive and exceeds $1/2$ when $|y| > 2$. Putting 5.17 and 5.18 together, we see that for every $\varepsilon > 0$ there is $K = [-2/b, 2/b]$ such that $\mu_n(K) \ge 1 - \varepsilon$ for all but finitely many n. By taking b smaller if necessary, we can ensure that $\mu_n(K) \ge 1 - \varepsilon$ for all n. Hence (μ_n) is tight.

Consequently, by Theorem 5.13, every subsequence N of $\mathbb{N}$ has a further subsequence N' such that (μ_n) converges weakly, along N', to some probability measure μ'. By the necessity part, then, (f_n) must converge, along N', to the Fourier transform of μ', and 5.16 implies that the Fourier transform of μ' is f, no matter what N' is. In other words, every subsequence of (μ_n) has a further subsequence that converges weakly to the same probability measure μ (whose Fourier transform is f). Hence (μ_n) converges weakly to μ. □

Convergence of characteristic functions

The corollary next is a partial translation of the preceding theorem. No proof is needed.

5.19 COROLLARY. *The sequence (X_n) converges in distribution to X if and only if*

$$\lim_n \mathbb{E}\ \exp\ irX_n = \mathbb{E}\ \exp\ irX\ , \qquad r \in \mathbb{R}.$$

Exercises and complements

5.20 *Uniqueness of Fourier transforms.* This is to provide a proof for Exercise II.2.35 using the tools from this section. Let μ and ν be probability measures on $\mathbb{R}$. If they have the same Fourier transform then they are equal. Show.

5.21 *Fatou's Lemma.* If $X_n \to X$ in distribution, then $\mathbb{E}\,|X| \le \liminf \mathbb{E}\,|X_n|$. Show this using the almost sure representation theorem.

5.22 *Sums.* Suppose that $X_n \to X$ and $Y_n \to Y$, both in distribution. Suppose that X_n and Y_n are independent for each n, and X and Y are independent. Show that $X_n + Y_n \to X + Y$ in distribution.

5.23 *Arithmetic of convergence in distribution.* If $X_n \to X$ in distribution then $a + bX_n \to a + bX$ in distribution for every fixed a in $\mathbb{R}$ and b in $\mathbb{R}$. Show.

5.24 *Continuation.* Let $X_n \to X$ and $Y_n \to y_0$, both in distribution, where y_0 is a fixed point. Then, $X_n + Y_n \to X + y_0$ and $X_n Y_n \to X y_0$, both in distribution. Show.

5.25 *Insensitivity.* If $X_n \to X$ in distribution and $\mathbb{P}\{X_n \ne Y_n\} \to 0$, then $Y_n \to X$ in distribution. Show.

5.26 *DeMoivre-Laplace Theorem.* Let $B_1, B_2, \ldots$ be independent Bernoulli variables with $\mathbb{P}\{B_n = 1\} = p$ and $\mathbb{P}\{B_n = 0\} = 1 - p$ for all n; here $p \in (0,1)$. Let $S_n = B_1 + \cdots + B_n$, and $Z_n = (S_n - np)/\sqrt{np(1-p)}$. Show that $Z_n \to Z$ in distribution, where Z has the standard Gaussian distribution.

5.27 *Convergence to Poisson distribution.* Recall that the Poisson distribution is a probability measure on $\mathbb{N} = \{0, 1, \ldots\}$ and $\mathbb{N}$ is a subset of $\mathbb{R}$. Let μ be the Poisson distribution on $\mathbb{R}$ with mean c, where c is a fixed number in $(0, \infty)$.

a) Show that (μ_n) converges weakly to μ if and only if $\mu_n(k-\varepsilon, k+\varepsilon) \to \mu\{k\}$ for every k in $\mathbb{N}$ and ε in $(0,1)$.

b) If the μ_n put all their mass on $\mathbb{N}$, then (μ_n) converges to μ weakly if and only if $\mu_n\{k\} \to \mu\{k\}$ for every k in $\mathbb{N}$.

5.28 *Convergence of binomial to Poisson.* For each integer $n \ge 1$, let $B_{n,1}, \ldots, B_{n,n}$ be independent Bernoulli variables with "success" probability p_n, that is, $p_n = \mathbb{P}\{B_{n,i} = 1\} = 1 - \mathbb{P}\{B_{n,i} = 0\}$. Let $S_n = B_{n,1} + \cdots + B_{n,n}$. Assuming that $\mathbb{E}\, S_n = np_n \to c$, show that the distribution of S_n converges weakly to the Poisson distribution with mean c.

5.29 *Weak convergence on metric spaces.* Let (E, d) be a metric space that is complete and separable. Let $\mathbb{C}_b = C_b(E, \mathbb{R})$ be the collection of all bounded continuous functions from E into $\mathbb{R}$. The σ-algebra on E is the Borel σ-algebra. Let $\mu_1, \mu_2, \ldots, \mu$ be probability measures on E, and $X_1, X_2, \ldots, X$ be E-valued random variables with those distributions.

a) Definition 5.1 makes sense within this setup and is the definition of weak convergence and convergence in distribution on metric spaces.

b) Characterization theorem 5.4 remains the same; even its proof remains good.

c) Comments on the equality of measures and uniqueness of weak limits remain in force.

d) Distribution functions don't make sense in the present setup. So, Proposition 5.7 is not meaningful. However, quantile functions are somewhat meaningful if we think of them in their generalized sense: there exists a measurable function $q : (0,1) \mapsto E$ such that $\mu = \lambda \circ q^{-1}$ where λ is the Lebesgue measure on $(0,1)$. Similarly for q_n and μ_n.

e) Theorem 5.8 remains true and is called the Skorokhod representation theorem. Its proof is, basically, still good but needs a lot more technical work in the definitions of $q_1, q_2, \ldots, q$. See the preceding remark. Of course, Corollary 5.9 remains true.

f) Proposition 5.10 is no longer applicable, since integrability is a concept for $\mathbb{R}$-valued variables (or $\mathbb{R}^d$-valued at most).

g) Proposition 5.11 is still good. So is its proof.

h) Definition 5.12 of tightness remains good, but the compact set K has to be compact in the metric space E now. Theorem 5.13 remains true, but its proof is no longer good.

i) Fourier transforms are not applicable to measures on E. But, if $E = \mathbb{R}^m$ with the usual distance, then Fourier transforms of probability measures are defined by

$$f(r) = \hat{\mu}(r) = \int_{\mathbb{R}^m} \mu(dx) e^{ir\cdot x} \ , \quad r \in \mathbb{R}^m \ ,$$

where $r \cdot x = \sum_1^m r_i x_i$, the inner product of r and x. With this definition, Theorem 5.15 remains good on $E = \mathbb{R}^m$, but the proof needs re-working. Of course, Corollary 5.19 is good.

6 Laws of Large Numbers

Our aim is to give an introduction to an important chapter of classical probability theory. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a probability space, $X_1, X_2, \ldots$ are real-valued random variables, and for $n \geq 1$,

6.1 $$S_n = X_1 + \cdots + X_n \ , \qquad \bar{X}_n = \frac{1}{n} S_n \ .$$

We start with the following classical result on the long run behavior of the averages $\bar{X}_n$. The statement about convergence in probability is called the *weak law of large numbers*, and the one about almost sure convergence is called the *strong law of large numbers.*

6.2 THEOREM. *Suppose that the X_n are pairwise independent and identically distributed with finite mean a and finite variance b. Then, $(\bar{X}_n)$ converges to a in L^2, in probability , and almost surely.*

Proof. i) Note that $\mathbb{E}\ S_n = na$, and Var $S_n = nb$ by the pairwise independence of the X_n. Thus, $\mathbb{E}\ \bar{X}_n = a$ and Var $\bar{X}_n = b/n$, and hence

$$\mathbb{E}\ |\bar{X}_n - a|^2 = \text{Var}\ \bar{X}_n \to 0 \quad \text{ as } n \to \infty.$$

In other words, $\bar{X}_n \to a$ in L^2. The convergence is in probability as well since it is implied by L^2-convergence.

ii) To show the almost sure convergence we start with the observation that it is enough to prove it assuming that the X_n are all positive. Because, then, the proof will apply to the sequences (X_n^+) and (X_n^-) separately. Assume $X_n \geq 0$ for all n. Let $N = (n_k)$ be defined by $n_k = k^2$, $k \in \mathbb{N}^*$. By Chebyshev's inequality,

$$\varepsilon^2 \sum_{n \in N} \mathbb{P}\left\{|\bar{X}_n - a| > \varepsilon\right\} \leq b \sum_{k=1}^{\infty} 1/k^2 < \infty$$

for every $\varepsilon > 0$. It follows from Proposition 2.6, a Borel-Cantelli lemma, that $\bar{X}_n \to a$ along N almost surely. Let Ω_0 be the almost sure set over which convergence holds.

For ω in Ω_0, since the numbers $X_n(\omega)$ are all positive, Lemma 1.7 is applicable to the sequence $(\bar{X}_n(\omega))$ with $r = \lim n_{k+1}/n_k = 1$. Thus, for every ω in Ω_0,

$$a \leq \liminf \bar{X}_n(\omega) \leq \limsup \bar{X}_n(\omega) \leq a \ ,$$

which completes the proof since Ω_0 is almost sure. □

Strong law of large numbers

In the preceding theorem, the assumption that the X_n have finite variance seems to be for reasons of simplifying the proof. Here, we shall remove that condition. We start with the extreme case where the expected value is $+\infty$. We limit this to positive sequences.

6.3 Proposition. *Suppose that the X_n are positive, pairwise independent, and identically distributed as a generic random variable X with $\mathbb{E}\ X = +\infty$. Then, $\bar{X}_n \to +\infty$ almost surely.*

Proof. Fix b in $\mathbb{R}_+$, let $Y_n = X_n \wedge b$, put $\bar{Y}_n = (Y_1 + \cdots + Y_n)/n$. Then, Theorem 6.2 applies to (Y_n) and shows that $\bar{Y}_n \to \mathbb{E}\ (X \wedge b)$ almost surely. Since $X_n \geq Y_n$ for all n, it follows that $\liminf \bar{X}_n \geq \lim \bar{Y}_n = \mathbb{E}(X \wedge b)$ almost surely. This is true for arbitrary b, and $\mathbb{E}(X \wedge b) \to \mathbb{E}X = +\infty$ as $b \to \infty$ by the monotone convergence theorem. Thus, $\liminf \bar{X}_n = +\infty$ almost surely. □

The following theorem seems to be the latest work. The clever proof is due to Etemadi.

6.4 Theorem. *Suppose that the X_n are pairwise independent and have the same distribution as a generic variable X. If $\mathbb{E}X$ exists (infinite values are allowed), then $\bar{X}_n \to \mathbb{E}X$ almost surely.*

Proof. Both the assumptions and the conclusion apply to positive sequences (X_n^+) and (X_n^-) separately. Thus, it is sufficient to prove the claim under the further assumption that the X_n and X are positive. In fact, by replacing X_n with $3 + X_n$, we may assume that $X_n \geq 3$. Then, the preceding proposition gives the proof if $\mathbb{E}X = \infty$. Hence, for the remainder of the proof, we assume that $3 \leq X < \infty$ and $\mathbb{E}X < \infty$.

i) We shall mostly work with

6.5 $$Y_n = X_n 1_{\{X_n < n\}} \ , \quad T_n = Y_1 + \cdots + Y_n \ , \quad \bar{Y}_n = T_n/n \ ,$$

because each Y_n is bounded and, overall, they do not differ much from the X_n: We have

$$\begin{aligned} \textstyle\sum_n \mathbb{P}\{X_n \neq Y_n\} &= \textstyle\sum_n \mathbb{P}\{X_n \geq n\} \\ &= \textstyle\sum_n \mathbb{P}\{X \geq n\} \leq \int_0^\infty dt\ \mathbb{P}\{X \geq t\} = \mathbb{E}X < \infty \ , \end{aligned}$$

which implies, through Borel-Cantelli lemma 2.5, that for almost every ω we have $X_n(\omega) = Y_n(\omega)$ for all but finitely many n. Therefore, it is sufficient to show that

6.6 $$\bar{Y}_n \to \mathbb{E}\ X \qquad \text{almost surely.}$$

ii) Being functions of pairwise independent random variables, the Y_n are pairwise independent, which allows us to write Var T_n as the sum of the variances of $Y_1, \ldots, Y_n$. Since the distribution of Y_n is the same as that of $X\ 1_{\{X<n\}}$,

6.7 $$\mathbb{E}\ T_n = \sum_1^n \mathbb{E}\ X\ 1_{\{X<i\}} = \mathbb{E}\ X \sum_{i>X} \delta_i[1,n] \ ,$$

6.8 $$\text{Var}\ T_n = \sum_1^n \text{Var}\ Y_i \leq \sum_1^n \mathbb{E}Y_i^2 = \mathbb{E}\ X^2 \sum_{i>X} \delta_i[1,n] \ ,$$

where δ_i is Dirac sitting at i as usual.

Let Z_n be the sum over i, so that $\mathbb{E}\ T_n = \mathbb{E}\ XZ_n$. Note that Z_n is the number of integers in the interval $(X, n]$. The sequence (XZ_n/n) is dominated by X and converges to X. Thus, by 6.7 and the dominated convergence theorem,

6.9 $$\mathbb{E}\ \bar{Y}_n = \mathbb{E}\ XZ_n/n \to \mathbb{E}\ X \ .$$

iii) Next, we show that 6.6 holds as $n \to \infty$ over the sequence $N = (n_k)$ defined by letting n_k be the smallest integer exceeding e^{ak}, where $a > 0$ is fixed. In view of 6.9, this is equivalent to showing that

6.10 $$\bar{Y}_n - \mathbb{E}\bar{Y}_n \to 0 \quad \text{almost surely along } N \ .$$

We do this by using a Borel-Cantelli lemma, Proposition 2.6, which requires that we show, for every $\varepsilon > 0$, that

6.11 $$s = \sum_{n \in N} \mathbb{P}\left\{|\bar{Y}_n - \mathbb{E}\bar{Y}_n| > \varepsilon\right\} < \infty.$$

We estimate the sum s using Chebyshev's inequality and 6.8:

$$\varepsilon^2 s \le \sum_{n\in N} \mathrm{Var}\ \bar{Y}_n = \sum_{n\in N} (\mathrm{Var}\ T_n)/n^2 \le \mathbb{E}\ X^2 \sum_{i>X} \sum_{k\ge m_i} (1/n_k)^2$$

where m_i is the smallest integer j with $n_j \ge i$. Recall that $n_k > e^{ak}$, note that $\exp am_i > i - 1$. It follows that the last sum over k is less than $c \cdot \exp(-2am_i) \le c/(i-1)^2$, where the constant c is $\sum_1^\infty e^{-2aj} = 1/(1-e^{-2a}) < \infty$. Thus, the sum over $i > X$ is less than the integral of c/x^2 over the interval $(X-2, \infty)$. So,

$$\varepsilon^2 s \le c\ \mathbb{E}\ X^2 \frac{1}{X-2} \le c\ \mathbb{E}(X+6) < \infty\ ,$$

where we used the assumptions that $X \ge 3$ and $\mathbb{E}X < \infty$. This shows 6.11, proves 6.10, and in view of 6.9, shows that

6.12 $$\bar{Y}_n \to \mathbb{E}\ X \quad \text{almost surely along } N.$$

iv) Let Ω_0 be the almost sure set of 6.12. For ω in Ω_0, since the $Y_n(\omega)$ are positive and $r = \lim n_{k+1}/n_k = e^a$, the subsequence lemma 1.7 applies to yield

$$e^{-a}\ \mathbb{E}X \le \liminf \bar{Y}_n(\omega) \le \limsup \bar{Y}_n(\omega) \le e^a\ \mathbb{E}X\ ,$$

which completes the proof of 6.6, and of the theorem, upon letting $a \to 0$. □

Weak law of large numbers

Returning to the classical weak law of Theorem 6.2, we observe that the proof of convergence in L^2 rests on two essential points: variance of S_n is the sum of the variances of $X_1, \ldots, X_n$ and $\mathrm{Var}\,(S_n/n) \to 0$. The first point holds if the X_n are pairwise uncorrelated, and the second point can be ensured by much weaker conditions on Var X_n. In fact, what seems needed is some increasing sequence (b_n) with $\mathrm{Var}\,(S_n/b_n) \to 0$. Here is an illustration. The proof is elementary given Kronecker's lemma 1.10.

6.13 Proposition. *Suppose that the X_n are uncorrelated and $\sum \mathrm{Var}(X_n/b_n) < \infty$ for some strictly positive sequence (b_n) increasing to $+\infty$. Then, $(S_n - \mathbb{E}\ S_n)/b_n \to 0$ in L^2 and in probability.*

Proof. Uncorrelatedness implies that

$$\mathbb{E}\left|\frac{S_n - \mathbb{E}\ S_n}{b_n}\right|^2 = \mathrm{Var}\ \frac{1}{b_n} S_n = (\frac{1}{b_n})^2 \sum_1^n \mathrm{Var}\ X_i\ .$$

Now, the assumption of summability for $\sum(\mathrm{Var}\ X_n)/b_n^2$ implies, through Kronecker's lemma 1.10, that the right side converges to 0. This proves the convergence in L^2, and the latter implies convergence in probability. □

In the preceding proposition, if the X_n are independent (instead of being merely uncorrelated), then the convergence is almost sure as well; see 7.10 below.

Exercises and complements

6.14 *Frequency interpretation of probabilities.* Let $B_1, B_2, \ldots$ be independent Bernoulli variables with success probability p, that is, $p = \mathbb{P}\{B_n = 1\} = 1 - \mathbb{P}\{B_n = 0\}$. Let $S_n = B_1 + \cdots + B_n$, the number of successes in the first n trials. Then, $\bar{B}_n = S_n/n$ is the success frequency during the first n trials. Show that $\bar{B}_n \to p$ almost surely. In the frequentist theory of probabilities, this is the way p is defined.

6.15 *Empirical distributions.* Let $X_1, X_2, \ldots$ be independent and identically distributed random variables taking values in some measurable space $(E, \mathcal{E})$. Let μ be their common distribution. Define

$$F_n(A) = \frac{1}{n} \sum_{k=1}^{n} 1_A \circ X_k \ , \quad A \in \mathcal{E} \ .$$

Then, for each ω in Ω, the mapping $A \mapsto F_n(\omega, A)$ is a probability measure on $(E, \mathcal{E})$; thus, F_n is called the *empirical distribution* corresponding to $X_1, \ldots, X_n$.

a) Show that, for each A in $\mathcal{E}$, $F_n(A) \to \mu(A)$ almost surely.

b) Show that, for every positive $\mathcal{E}$-measurable f,

$$\frac{1}{n} \sum_{k=1}^{n} f \circ X_k \to \mu f$$

almost surely. This result is used to estimate μf by Monte Carlo methods.

6.16 *Glivenko-Cantelli.* This is a continuation of the preceding, but with $E = \mathbb{R}$. Let $c_n(\omega, x) = F_n(\omega, (-\infty, x])$ and $c(x) = \mu(-\infty, x]$. The preceding exercise shows that, for each x, $c_n(\omega, x) \to c(x)$ for almost every ω. In fact, as shown by Glivenko and Cantelli, the convergence is uniform in x, that is,

$$\sup{}_{x \in \mathbb{R}} |c_n(\omega, x) - c(x)| \to 0$$

for almost every ω.

a) It can be shown that $\sup_x |c_n(x) - c(x)| \to 0$ provided that $c_n(r) \to c(r)$ for each rational point r and $c_n(x) - c_n(x-) \to c(x) - c(x-)$ at each point x of discontinuity for c.

b) We apply the preceding to $c_n(\omega, x)$ and $c(x)$: For each rational r, there is an almost sure event Ω_r such that $c_n(\omega, r) \to c(r)$ for every ω in Ω_r; this is by 6.15 applied with $A = (-\infty, r]$. For each discontinuity point x for c, there is an almost sure event Ω'_x such that $c_n(\omega, x) - c_n(\omega, x-) \to c(x) - c(x-)$ for every ω in Ω'_x; this is by 6.15 applied with $A = \{x\}$. Since the rationals are countable, $\Omega_0 = \cup_r \Omega_r$ is an almost sure event. Since a distribution function has at most countably many points of discontinuity, the set $\Omega'_0 = \cup_x \Omega'_x$ is an almost sure event. Thus, if ω belongs to the almost sure event $\Omega_0 \cup \Omega'_0$, then part (a) applies to show that $c_n(\omega, x) \to c(x)$ uniformly in x.

6.17 *Inversion of Laplace transforms of functions.* Let f be a bounded continuous function on $\mathbb{R}_+$, and let g be its Laplace transform, that is,

$$g(r) = \int_{\mathbb{R}_+} dt\; e^{-rt} f(t) \;, \quad r \in \mathbb{R}_+ \;.$$

Then,

$$f(t) = \lim_{n\to\infty} \frac{(-1)^n}{n!} (\frac{n}{t})^{n+1} g^{(n)}(\frac{n}{t}) \;, \quad t \in \mathbb{R}_+ \;,$$

where $g^{(n)}$ is the nth derivative of g. Show this by using Theorem 6.2, the classical strong law of large numbers. Hint: Reduce the claim to showing that

$$f(t) = \lim_{n\to\infty} \mathbb{E}\, f(tS_{n+1}/n)$$

where $S_{n+1} = X_1 + \cdots + X_{n+1}$, and the X_i are independent and identically distributed exponential variables with mean 1.

6.18 *Laplace transforms of measures.* Let μ be a probability measure on $\mathbb{R}_+$ and let

$$g(r) = \int_{\mathbb{R}_+} \mu(dx) e^{-rx} \;, \quad r \in \mathbb{R}_+ \;.$$

Then, for each t for which $\mu\{t\} = 0$,

$$\lim_{n\to\infty} \sum_{k \le nt} \frac{(-n)^k}{k!} g^{(k)}(n) = \mu[0,t] \;;$$

here, $g^{(k)}$ is the kth derivative of g. Show this by following the steps below.

a) Let $X_1, X_2, \ldots$ be independent Poisson distributed with mean x. Then, $S_n = X_1 + \cdots + X_n$ has the Poisson distribution with mean nx. Put $\bar{X}_n = S_n/n$. Show that

$$\lim_{n\to\infty} \mathbb{P}\{\bar{X}_n \le t\} = \begin{cases} 1 & \text{if } x < t \;, \\ 0 & \text{if } x > t \;. \end{cases}$$

b) Show that

$$\mathbb{P}\{\bar{X}_n \le t\} = \sum_{k \le nt} \frac{e^{-nx}(nx)^k}{k!} := f_n(x,t).$$

c) Show that the claim is the convergence of

$$\int_{\mathbb{R}_+} \mu(dx) f_n(x,t)$$

to $\mu[0,t]$ for every t that is not an atom for μ. Complete the proof.

7 Convergence of Series

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let (X_n) be a sequence of real-valued random variables. We are interested in the almost sure convergence of the series $\sum X_n$, in other words, the almost sure convergence of the sequence (S_n), where $S_n = X_1 + \cdots + X_n$. All the results below are for the case where the X_n are independent, in which case Kolmogorov's 0-1 law aplies, and the convergence of the series has probability 0 or 1, the better case being our aim.

Inequalities for maxima

Suppose that the X_n have mean 0. Then, Chebyshev's inequality yields

$$\varepsilon^2 \ \mathbb{P}\{|S_n| > \varepsilon\} \le \text{Var } S_n = \mathbb{E} \ S_n^2 \ .$$

The following is a considerable improvement when the X_n are independent; it is called *Kolmogorov's inequality.*

7.1 Lemma. *Suppose that the X_n are independent and have mean 0. Then, for every a in $(0, \infty)$,*

$$a^2 \ \mathbb{P}\{\max_{k \le n} |S_k| > a\} \le \text{Var } S_n \ .$$

Proof. Fix $a > 0$ and $n \ge 1$. Define $N(\omega) = \inf\{k \ge 1 : |S_k(\omega)| > a\}$ for every ω in Ω. Note that $N(\omega) = k$ if and only if $|S_k(\omega| > a$ and $|S_j(\omega)| \le a$ for all $j < k$. Thus $1_{\{N=k\}}$ is a function of $(X_1, \ldots, X_k)$, which shows that N is a random variable. Moreover, by the same reason, for $k < n$, $U = S_k \ 1_{\{N=k\}}$ and $V = S_n - S_k$ are functions of independent vectors $(X_1, \ldots, X_k)$ and $(X_{k+1}, \ldots, X_n)$, and thus $\mathbb{E} \ UV = \mathbb{E} \ U \ \mathbb{E} \ V$; and $\mathbb{E} \ V = 0$ since $\mathbb{E} \ X_i = 0$ for all i by assumption. Hence, for $k \le n$,

7.2 $$\mathbb{E} \ S_k(S_n - S_k) \ 1_{\{N=k\}} = 0 \ .$$

Note that $S_n^2 = [S_k + (S_n - S_k)]^2 \ge S_k^2 + 2S_k(S_n - S_k)$, and that $|S_k|^2 > a^2$ on the event $\{N = k\}$. Thus

$$\mathbb{E} \ S_n^2 \ 1_{\{N=k\}} \ge a^2 \ \mathbb{E} \ 1_{\{N=k\}} + 2 \ \mathbb{E} \ S_k(S_n - S_k) 1_{\{N=k\}} = a^2 \ \mathbb{P}\{N = k\} \ ,$$

in view of 7.2. Summing both sides over $k \le n$ and reversing the order, we get

$$a^2 \ \mathbb{P}\{N \le n\} \le \mathbb{E} \ S_n^2 1_{\{N \le n\}} \le \mathbb{E} \ S_n^2 = \text{Var } S_n \ ,$$

which completes the proof upon noting that the event $\{N \le n\}$ is the same as the event that $\{\max_{k \le n} |S_k| > a\}$. □

The assumption of independence for the X_n will be relaxed later by martingaling. For the present, the following is an estimate going in the opposite direction.

7.3 Lemma. *Suppose that the X_n are independent, have mean zero, and are dominated by some constant b. Then, for every $a > 0$,*

$$\mathbb{P}\{\max_{k\le n}|S_k| > a\} \ge 1 - (a+b)^2/\text{Var } S_n \ .$$

Proof. Fix n and a. Let N be as in the preceding proof. Now the claim is that

7.4
$$\mathbb{P}\{N > n\}\text{Var } S_n \le (a+b)^2 \ .$$

Fix $k \le n$. Write $S_n^2 = S_k^2 + 2S_k(S_n - S_k) + (S_n = S_k)^2$; and note that $|S_k(\omega)| \le a+b$ if $N(\omega) = k$, because $|S_{k-1}(\omega)| \le a$ by the definition of $N(\omega)$ and $|X_k(\omega)| \le b$ by the assumed boundedness. Thus,

$$S_n^2 1_{\{N=k\}} \le (a+b)^2 1_{\{N=k\}} + 2S_k(S_n - S_k)1_{\{N=k\}} + (S_n - S_k)^2 1_{\{N=k\}} \ .$$

On the right side, the expectation of the second term is 0 by 7.2, and the reasoning leading to 7.2 shows that the expectation of the third term is $\mathbb{E}(S_n - S_k)^2\mathbb{P}\{N = k\} \le \mathbb{P}\{N = k\}\text{Var}S_n$. Hence, taking expectations on both sides and adding over $k \le n$,

$$\mathbb{E}\ S_n^2 1_{\{N\le n\}} \le [(a+b)^2 + \text{Var}S_n]\mathbb{P}\{N \le n\} \ .$$

On the other hand, for every ω, if $N(\omega) > n$ then $|S_n(\omega)| \le a$. So,

$$\mathbb{E}\ S_n^2 1_{\{N>n\}} \le \mathbb{E}\ a^2 1_{\{N>n\}} = a^2\mathbb{P}\{N > n\} \ .$$

Adding the last two expressions side by side we get an upper bound for $\text{Var}S_n = \mathbb{E}S_n^2$; and rearranging the terms somewhat we obtain 7.4. □

Convergence of series and variances

The following shows that the summability of variances implies the convergence of the associated series:

7.5 Theorem. *Suppose that the X_n are independent and have zero mean. If $\sum \text{Var } X_n$ converges then $\sum X_n$ converges almost surely.*

Proof. By Kolmogorov's inequality applied to the sequence $(X_{n+m})_{m\ge1}$, for every $\varepsilon > 0$,

$$\varepsilon^2\mathbb{P}\{\max{}_{k\le m}|S_{n+k} - S_n| > \varepsilon\} \le \sum_{n+1}^{n+m} \text{Var } X_i \ .$$

Assume that $\sum \text{Var } X_n < \infty$. Then, the right side goes to 0 as we let $m \to \infty$ first and $n \to \infty$ next. Hence, the condition of Proposition 2.14 is satisfied, and (S_n) converges almost surely. □

The following is nearly a converse to the preceding theorem; within the proof, the interesting trick in the second step is called symmetrization.

7.6 PROPOSITION. *Suppose that (X_n) is a bounded sequence of independent variables. If $\sum(X_n - a_n)$ is almost surely convergent for some sequence (a_n) in $\mathbb{R}$, then $\sum \text{Var} X_n < \infty$.*

Proof. i) First, we prove the assertion under the extra conditions that $a_n = 0$ and $\mathbb{E}\, X_n = 0$ for all n. Let b be a bound for (X_m). Note that

$$Z_m = \sup_k |S_{m+k} - S_m| = \lim_n \max_{k \le n} |S_{m+k} - S_m|,$$

and the limit is of an increasing sequence. Thus, for every $\varepsilon > 0$,

$$7.7 \quad \mathbb{P}\{Z_m > \varepsilon\} = \lim_n \mathbb{P}\{\max_{k \le n} |S_{m+k} - S_m| > \varepsilon\} \ge 1 - \frac{(\varepsilon + b)^2}{\sum_{m+1}^{\infty} \text{Var } X_i} \ ,$$

where we used Lemma 7.3 applied to the sequence $(X_{m+n})_{n \ge 1}$. If (S_n) converges almost surely, then $Z_m \to 0$ almost surely by Lemma 2.13, and thus the left side of 7.7 tends to 0 as $m \to \infty$. This is impossible if $\sum \text{Var } X_i = +\infty$.

ii) Next we remove the extra conditions. Let (Y_n) be independent of (X_n) and have the same law. Suppose that $\sum(X_n - a_n)$ is almost surely convergent. Then, so is $\sum(Y_n - a_n)$ since the sequences (X_n) and (Y_n) have the same law. Thus, $\sum(X_n - Y_n) = \sum(X_n - a_n) - \sum(Y_n - a_n)$ converges almost surely and the sequence $(X_n - Y_n)_{n \ge 1}$ is bounded and $\mathbb{E}(X_n - Y_n) = 0$ for all n. Hence, part (i) of the proof applies, and we must have $\sum \text{Var}(X_n - Y_n) < \infty$. This completes the proof since $\text{Var}(X_n - Y_n) = 2\text{Var}(X_n)$. □

Kolmogorov's three series theorem

This theorem gives necessary and sufficient conditions for the almost sure convergence of the series $\sum X_n$. Essentially, it amounts to combining Theorem 7.5 and Proposition 7.6. Since (X_n) is generally not bounded, the conditions are expressed in terms of the *truncated* variables

$$7.8 \qquad Y_n = X_n 1_{\{|X_n| \le b\}} \ ,$$

where b is a fixed constant in $(0, \infty)$.

7.9 THEOREM. *Suppose that the X_n are independent. Then, $\sum X_n$ is almost surely convergent if and only if the following three series are convergent:*

$$7.10 \qquad \sum \mathbb{P}\{X_n \ne Y_n\} \ , \quad \sum \mathbb{E}\, Y_n \ , \quad \sum \text{Var } Y_n \ .$$

Proof. *Sufficiency.* Suppose that all three series in 7.10 are convergent. The independence of the X_n imply the independence of the Y_n. The convergence of the third series implies, via Theorem 7.5, that $\sum(Y_n - \mathbb{E}Y_n)$ converges almost surely. This and the convergence of the second series together imply that $\sum Y_n$ converges almost surely. The last implies that $\sum X_n$ converges almost surely, because the convergence of the first series shows, via Borel-Cantelli 2.5, that for almost every ω the numbers $X_n(\omega)$ and $Y_n(\omega)$ differ for at most finitely many n.

Necessity. Suppose that $\sum X_n$ is convergent almost surely. Then, for almost every ω, there are at most finitely many n with $|X_n(\omega)| > b$, which means that $X_n(\omega) \neq Y_n(\omega)$ for only finitely many n. Thus, $\sum 1_{\{X_n \neq Y_n\}} < \infty$ almost surely, and the independence of the X_n implies that the events $\{X_n \neq Y_n\}$ are independent. It follows from Borel-Cantelli lemma's divergence part, Proposition 2.9b, that the first series in 7.10 must converge.

Consequently, $\sum Y_n$ is almost surely convergent (since $\sum X_n$ is so). Now, Proposition 7.6 implies that the third series in 7.10 converges. This in turn implies via Theorem 7.5 that $\sum(Y_n - \mathbb{E}Y_n)$ is almost surely convergent, which together with the convergence of $\sum Y_n$ imply that the second series in 7.10 is convergent. □

Application to strong laws

This is to show that Proposition 6.14 regarding the weak law can be altered to a strong law. In fact, the method here is the classical way of proving the strong law of large numbers.

7.11 Proposition. *Suppose that the X_n are independent and $\sum \text{Var}(X_n/b_n) < \infty$ for some strictly positive sequence (b_n) increasing to $+\infty$. Then, $(S_n - \mathbb{E}S_n)/b_n \to 0$ almost surely (as well as in L^2).*

Proof. The condition implies, through Theorem 7.5, that $\sum(X_n - \mathbb{E}X_n)/b_n$ converges almost surely. In turn, this implies, through Kronecker's lemma 1.7, the desired conclusion. □

8 Central Limits

This section is a short introduction to a topic of central importance to classical probability theory: convergence of the distributions of sums of independent random variables to Gaussian, Poisson, and other infinitely divisible distributions. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a probability space in the background. Also, we let $\mathfrak{Z}$ denote a generic random variable with the standard Gaussian distribution (with mean 0 and variance 1), and $\mathfrak{P}_c$ a generic variable with Poisson distribution with mean c.

We start with the following generalization of DeMoivre-Laplace theorem; see Exercise 5.26.

8.1 THEOREM. *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean a and variance b, both finite. Let $S_n = X_1 + \cdots + X_n$ and $Z_n = (S_n - na)/\sqrt{nb}$, $n \geq 1$. Then, (Z_n) converges to $\mathfrak{Z}$ in distribution.*

REMARK. The claim is that (see Proposition 5.7)

$$\lim_{n\to\infty} \mathbb{P}\left\{ \frac{S_n - na}{\sqrt{nb}} \leq x \right\} = \int_{-\infty}^{x} dy \frac{1}{\sqrt{2\pi}} e^{-y^2/2} , \quad x \in \mathbb{R} ,$$

which is the usual way of presenting results like this.

Proof. Let f denote the characteristic function of $(X_n - a)/\sqrt{b}$, which variable has mean 0 and variance 1. Thus, a convenient version of Taylor's theorem yields

$$f(r) = f(0) + f'(0)r + \frac{1}{2} f''(0) r^2 (1 + h(r)) = 1 - \frac{1}{2} r^2 (1 + h(r))$$

for some function h with $|h(r)| \to 0$ as $r \to 0$. Since the X_n are independent,

$$\mathbb{E} \exp irZ_n = \left[f(\frac{r}{\sqrt{n}}) \right]^n = \left[1 - \frac{r^2/2}{n} (1 + h(\frac{r}{\sqrt{n}})) \right]^n \to e^{-r^2/2}$$

as $n \to \infty$, since $(1 + c_n/n)^n \to e^c$ if c is the limit of the complex numbers c_n. Noting that $\exp(-r^2/2)$ is the characteristic function of $\mathfrak{Z}$ completes the proof via Corollary 5.19. □

The preceding is the most famous of the central limit theorems. Assuming $a = 0$ and $b = 1$, which is without loss of generality, the essential idea is the following: For large n, the variable $Z_n = S_n/\sqrt{n}$ is the sum of "small" independent random quantities

$$X_{n,1} = \frac{1}{\sqrt{n}} X_1, \; \ldots \; , X_{n,n} = \frac{1}{\sqrt{n}} X_n \; .$$

As $n \to \infty$, these summands approach 0 but their sum Z_n has mean 0 and variance 1 for all n. Unfortunately, such ideas got hidden under the analytic machinery of the preceding proof. The re-formulation below and the proofs to follow are more illuminating.

Triangular arrays

For the remainder of this section, we shall deal with an infinite double-array

8.2 $$[X_{nj}] = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots \\ X_{21} & X_{22} & X_{23} & \cdots \\ X_{31} & X_{32} & X_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

of real-valued random variables. For each n, there is an integer k_n such that $X_{nj} = 0$ for all $j > k_n$. The sequence (k_n) is increasing with limit $+\infty$, and therefore the array above is basically triangular. We let Z_n denote the nth row sum:

8.3 $$Z_n = \sum_j X_{nj} \; ;$$

here (and below in similar sums) the sum is over all j; the effective range is $\{1, 2, \ldots, k_n\}$, but k_n will not appear explicitly.

Throughout, for each n, it is assumed that the variables on the nth row are independent. But those on different rows may depend on each other; in fact, in the classical case, the $(n+1)^{\text{th}}$ row determines the nth.

Liapunov's Theorem

This is formulated with a condition on the third moments of the X_{nj}.

8.4 THEOREM. *Suppose that* $\mathbb{E}\, X_{nj} = 0$ *for all* n *and* j*, and* $\operatorname{Var} Z_n = 1$ *for all* n*, and* $\lim_n \sum_j \mathbb{E}|X_{nj}|^3 = 0$*. Then,* $Z_n \to \mathfrak{Z}$ *in distribution.*

Proof. We put the essential part as a lemma below. Applying the lemma to cos rx and sin rx separately, we get

$$|\mathbb{E}\; \exp\; irZ_n - \mathbb{E}\; \exp\; ir\mathfrak{Z}| \le 2|r|^3 \textstyle\sum_j \mathbb{E}|X_{nj}|^3 \; .$$

The conclusion is immediate from Corollary 5.19. □

The lemma needed, due to Lindeberg, considers a sum of k independent variables and approximates the distribution of the sum by the Gaussian distribution with the same mean and variance as the sum.

8.5 LEMMA. *Let* $Y_1, \ldots, Y_k$ *be independent and have mean 0. Let* S *be their sum and assume that* $\operatorname{Var} S = 1$*. Let* $f : \mathbb{R} \mapsto \mathbb{R}$ *be differentiable thrice and assume that the derivatives* f', f'', f''' *are bounded and continuous, with* c *a bound for* $|f'''|$*. Then,*

$$|\mathbb{E}\; f \circ S - \mathbb{E}\; f \circ \mathfrak{Z}| \le c \sum_1^k \mathbb{E}\; |Y_j|^3 \; .$$

Proof. Let $Z_1, \ldots, Z_k$ be independent Gaussian variables with means $\mathbb{E}\; Z_j = \mathbb{E}\; Y_j = 0$ and variances $\operatorname{Var} Z_j = \operatorname{Var} Y_j$. Then $T = Z_1 + \cdots + Z_k$ has the same distribution as $\mathfrak{Z}$ and the claim is that

8.6 $$|\mathbb{E}\; f \circ S - \mathbb{E}\; f \circ T| \le c \sum_1^k \mathbb{E}\; |Y_j|^3 \; .$$

The idea is to replace, one at a time, each Y_j with Z_j. So, we define $V_1, \ldots, V_k$ recursively by

$$S = V_1 + Y_1 \; ; \quad V_j + Z_j = V_{j+1} + Y_{j+1} \; , \quad 1 \le j < k \; ,$$

and note that $V_k + Z_k = T$. Then,

$$f \circ S - f \circ T = \sum_1^k [f(V_j + Y_j) - f(V_j + Z_j)] \ ,$$

and to prove 8.6, it is enought to show that, for each j,

8.7 $$|\mathbb{E}\ f(V_j + Y_j) - \mathbb{E}\ f(V_j + Z_j)| \le c\mathbb{E}|Y_j|^3 \ .$$

To this end, fix j and drop it from notation. Start from Taylor's expansion for f:

8.8 $$f(v + x) = f(v) + f'(v)x + \frac{1}{2} f''(v)x^2 + \frac{1}{6} R(v)x^3$$

where $|R(v)| \le c$, the bound for $|f'''|$. Note that V, Y, Z are independent, and f, f', f'' are bounded, and $\mathbb{E}\ Y = \mathbb{E}\ Z = 0$, and $\mathbb{E}\ Y^2 = \mathbb{E}\ Z^2 = b^2$, say ($b$ does depend on j). Now, replace v with V and x with Y and take expectations on both sides of 8.8; then, replace v with V and x with Z and take expectations on both sides of 8.8; taking differences we get

8.9 $$|\mathbb{E}\ f(V + Y) - \mathbb{E}\ f(V + Z)| \le \frac{c}{6}(\mathbb{E}\ |Y|^3 + \mathbb{E}\ |Z|^3) \ .$$

Since Z has the Gaussian distribution with mean 0 and variance $\mathbb{E}\ Z^2 = \mathbb{E}\ Y^2 = b^2$, a direct computation shows that $\mathbb{E}\ |Z|^3 = b^3\sqrt{8/\pi} \le 2b^3$. Since L^2-norm is at most equal to L^3-norm, $b = (\mathbb{E}\ Y^2)^{1/2} \le (\mathbb{E}|Y|^3)^{1/3}$. Hence,

$$\mathbb{E}\ |Y|^3 + \mathbb{E}\ |Z|^3 \le \mathbb{E}\ |Y|^3 + 2\mathbb{E}\ |Y|^3 = 3\mathbb{E}\ |Y|^3 \ .$$

Putting this into 8.9 shows 8.7 and completes the proof. □

8.10 Remark. Going back to Liapunov's theorem, we note that the norming hypotheses on the means and variances are harmless. Suppose that Z_n has mean a_n and variance b_n^2. Then, the theorem applies to the triangular array $[Y_{nj}]$ with $Y_{nj} = (X_{nj} - \mathbb{E}\ X_{nj})/b_n$ to show that, if $\lim b_n^{-3} \sum_j \mathbb{E}\ |X_{nj} - \mathbb{E}\ X_{nj}|^3 = 0$, then $(Z_n - a_n)/b_n \to \mathfrak{Z}$ in distribution.

8.11 Corollary. *Let $a_n = \mathbb{E}\ Z_n$ and $b_n^2 = \operatorname{Var} Z_n$. Suppose that $a_n \to a$ and $b_n \to b$, where $b \ne 0$. Assume that for each n and j there is a constant c_{nj} such that $|X_{nj}| \le c_{nj}$ and that $\lim_n \sup_j c_{nj} = 0$. Then, $(Z_n - a)/b \to \mathfrak{Z}$ in distribution.*

Proof. Put $Y_{nj} = (X_{nj} - \mathbb{E}\ X_{nj})/b_n$. Note that $|Y_{nj} \le 2c_{nj}/b_n \le \varepsilon_n$, where $\varepsilon_n = 2(\max_j c_{nj})/b_n$, and therefore $|Y_{nj}|^3 \le \varepsilon_n |Y_{nj}|^2$. Thus,

$$\sum_j \mathbb{E}\ |Y_{nj}|^3 \le \varepsilon_n \sum_j (\operatorname{Var} X_{nj})/b_n = \varepsilon_n \ ,$$

and $\varepsilon_n \to 0$ in view of the assumptions on b_n and c_{nj}. Hence, Theorem 8.4 applies to the array $[Y_{nj}]$ to show that $(Z_n - a_n)/b_n = \sum_j Y_{nj} \to \mathfrak{Z}$ in distribution. Since $a_n \to a$ and $b_n \to b$, this implies (via Exercise 5.24) that $(Z_n - a)/b \to \mathfrak{Z}$ in distribution. □

Lindeberg's Theorem

The idea here is to replace the condition on third moments by something weaker, called *Lindeberg's condition*: For every $\varepsilon > 0$,

8.12 $$L_n(\varepsilon) = \sum_j \mathbb{E}\ X_{nj}^2\ i_\varepsilon \circ |X_{nj}| \to 0, \quad n \to \infty\ ,$$

this being after the reduction assumptions that $\mathbb{E}\ X_{nj} = 0$ and $\operatorname{Var} Z_n = 1$, and with the notation $i_\varepsilon = 1_{(\varepsilon,\infty)}$ as before and $\bar{i}_\varepsilon = 1 - i_\varepsilon = 1_{[0,\varepsilon]}$ to come.

8.13 THEOREM. *Suppose that* $\mathbb{E}\ X_{nj} = 0$ *and* $\operatorname{Var} Z_n = 1$ *for all* n *and* j. *If 8.12 holds then* $Z_n \to \mathfrak{Z}$ *in distribution.*

Proof. Assume 8.12. Then, for every $\varepsilon > 0$ there is an integer $m(\varepsilon)$ such that $L_n(\varepsilon) \le \varepsilon^3$ for all $n \ge m(\varepsilon)$, and we may choose $m(\varepsilon)$ to be increasing to $+\infty$ as ε decreases to 0. Choose ε_n for each n so that $m(\varepsilon_n) \le n$ for all n large enough (essentially, $n \mapsto \varepsilon_n$ is the functional inverse of $\varepsilon \mapsto m(\varepsilon)$). Then

8.14 $$\lim_n \left(\frac{1}{\varepsilon_n}\right)^2 L_n(\varepsilon_n) = 0\ .$$

Let $Y_{nj} = X_{nj} \bar{i}_{\varepsilon_n} \circ |X_{nj}|$ and put $S_n = \sum_j Y_{nj}$. Then,

$$\mathbb{P}\{Z_n \ne S_n\} \le \sum_j \mathbb{P}\{X_{nj} \ne Y_{nj}\} = \sum_j \mathbb{P}\{|X_{nj}| > \varepsilon_n\} \le \left(\frac{1}{\varepsilon_n}\right)^2 L_n(\varepsilon_n)$$

where the last inequality follows from the observation that $\varepsilon^2 i_\varepsilon \circ |X| \le X^2 i_\varepsilon \circ |X|$ for arbitrary ε and X. As $n \to \infty$, the right-most member goes to 0; thus, Exercise 5.25 applies, and to complete the proof, it is enough to show that $S_n \to \mathfrak{Z}$ in distribution. But, since $|Y_{nj}| \le \varepsilon_n$ and $\varepsilon_n \to 0$, Corollary 8.11 above implies that $S_n \to \mathfrak{Z}$ in distribution once we show that

8.15 $$\mathbb{E}\ S_n \to 0\ , \quad \operatorname{Var} S_n \to 1\ .$$

To show 8.15, we estimate the mean and variance of Y_{nj} with all subscripts dropped. Since $\mathbb{E}\ X = 0$, we have $\mathbb{E}\ Y = \mathbb{E}\ Y - \mathbb{E}\ X = -\mathbb{E}\ X i_\varepsilon \circ |X|$, which yields

8.16 $$|\mathbb{E}\ Y| \le \mathbb{E}\ |X| i_\varepsilon \circ |X| \le \frac{1}{\varepsilon} \mathbb{E}\ X^2 i_\varepsilon \circ |X|\ .$$

Second, using the first inequality of 8.16

8.17 $$\begin{aligned} \operatorname{Var} Y &\ge \mathbb{E}\ X^2 \bar{i}_\varepsilon \circ |X| - (\mathbb{E}\ |X| i_\varepsilon \circ |X|)^2 \\ &\ge \mathbb{E}\ X^2 \bar{i}_\varepsilon \circ |X| - \mathbb{E}\ X^2 i_\varepsilon \circ |X| = \mathbb{E}\ X^2 - 2\mathbb{E}\ X^2 i_\varepsilon \circ |X|\ , \end{aligned}$$

where we used Jensen's inequality to justify the second inequality. Third, in the other direction,

8.18 $$\operatorname{Var} X = \mathbb{E}\ X^2 \ge \mathbb{E}\ Y^2 \ge \operatorname{Var} Y\ .$$

Finally, we put back the subscripts n and j in 8.16, 8.17, 8.18, and sum them over j to get (recall that $Var\ Z_n = 1$)

8.19
$$|\mathbb{E}\ S_n| \le \frac{1}{\varepsilon_n} L_n(\varepsilon_n)\ , \quad 1 - 2L_n(\varepsilon_n) \le \text{Var}\ S_n \le 1\ .$$

Now 8.14 and 8.19 together imply 8.15, and the proof is complete. □

8.20 Remarks. a) In addition to ensuring that $Z_n \to \mathfrak{Z}$ in distribution, Lindeberg's condition implies that the X_{nj} are uniformly small compared with Z_n for large n, that is,

8.21
$$\lim_n \mathbb{P}\{\max{}_j |X_{nj}| > \varepsilon\} = 0 \qquad \text{for every } \varepsilon > 0\ .$$

To see this, let H_n be the event here, note that $H_n = \cup_j \{|X_{nj}| > \varepsilon\}$, and use Boole's inequality and a Chebyshev type argument to get $\mathbb{P}(H_n) \le L_n(\varepsilon)/\varepsilon^2$.

b) It follows that Lindeberg's condition is not necessary for $Z_n \to \mathfrak{Z}$ in distribution. For example, assuming $\mathbb{E}\ Z_n = 0$ and $\text{Var}\ Z_n = 1$ as before, suppose that all the X_{nj} have Gaussian distributions and, in particular, $\text{Var}\ X_{n,1} = 1/2$ for all n. In this case, Z_n has the same standard Gaussian distribution as $\mathfrak{Z}$, and $Z_n \to \mathfrak{Z}$ trivially, but 8.21 fails because of the chunk $X_{n,1}$ that does not get small with n.

c) However, assuming $\mathbb{E}\ X_{nj} = 0$ and $\text{Var}\ Z_n = 1$ as before, Lindeberg's condition is necessary and sufficient in order that $Z_n \to \mathfrak{Z}$ in distribution *and* that

8.22
$$\lim_n \max_j \mathbb{P}\{|X_{nj}| > \varepsilon\} = 0 \quad \text{for every } \varepsilon > 0\ .$$

Its sufficiency is immediate from the preceding theorem and the Remark (a) above, since 8.21 implies 8.22. The proof of its necessity requires hard work.

Feller-Lévy theorem

The essential assumptions in Lindeberg's theorem concern the finiteness of second moments. In the most general case, where "nothing is assumed," we have the following theorem. We omit the proof.

8.23 Theorem. *In order that 8.22 hold and $(Z_n - a_n)$ converges in distribution to $\mathfrak{Z}$ for some sequence (a_n) in $\mathbb{R}$, it is necessary and sufficient that, for every $\varepsilon > 0$*

$$\lim_n \sum_j \mathbb{E}\ |X_{nj}| i_\varepsilon \circ |X_{nj}| = 0\ , \quad \lim_n \sum_j \text{Var}\ X_{nj} \bar{i}_\varepsilon \circ |X_{nj}| = 1\ .$$

Convergence to Poisson distribution

Consider the array 8.2 with the assumptions stated there. Suppose that the X_{nj} take values in $\mathbb{N} = \{0, 1, \ldots\}$. Then, row sums Z_n take values in $\mathbb{N}$

and, without further norming, if (Z_n) converges in distribution, the limit must take values in $\mathbb{N}$. The following is about convergence to Poisson variables. We let $\mathfrak{P}_c$ denote the generic random variable having the Poisson distribution with mean c, where c is a constant in $(0, \infty)$.

8.24 THEOREM. *Let the X_{nj} take values in $\mathbb{N}$. Suppose that*

$$\lim_n \max_j \mathbb{P}\{X_{nj} \geq 1\} = 0\ , \ \lim_n \sum_j \mathbb{P}\{X_{nj} \geq 1\} = c\ , \ \lim_n \sum_j \mathbb{P}\{X_{nj} \geq 2\} = 0\ .$$

Then, $Z_n \to \mathfrak{P}_c$ in distribution.

Proof. Put $B_{nj} = X_{nj} \wedge 1$ and $S_n = \sum_j B_{nj}$. Then,

$$\mathbb{P}\{Z_n \neq S_n\} \leq \sum_j \mathbb{P}\{X_{nj} \neq B_{nj}\} = \sum_j \mathbb{P}\{X_{nj} \geq 2\} \to 0$$

as $n \to \infty$. Thus, by Exercise 5.25, it is enough to show that $S_n \to \mathfrak{P}_c$. In turn, the latter is equivalent to showing that the Laplace transforms converge correctly; that is, we need to show that

$$\mathbb{E}\ \exp(-rS_n) = \prod_j \left[\mathbb{P}\{B_{nj} = 0\} + e^{-r}\mathbb{P}\{B_{nj} = 1\}\right] \to \exp\left[-c(1 - e^{-r})\right]$$

for every r in $\mathbb{R}_+$. To this purpose, fix r, let $x_{nj} = (1 - e^{-r})\mathbb{P}\{Y_{nj} = 1\}$, and put $b = (1 - e^{-r})c$. With these notation, what we need is translated into proving the next lemma. □

8.25 LEMMA. *Let x_{nj} form a triangular (that is, for each n there is k_n such that $x_{nj} = 0$ for all $j > k_n$) array of positive numbers. Suppose that*

$$\lim_n \max_j x_{nj} = 0\ , \quad \lim_n \sum_j x_{nj} = b\ .$$

Then, $\lim_n \prod_j (1 - x_{nj}) = e^{-b}$.

Proof. Since $\lim \max x_{nj} = 0$, we have $x_{nj} \leq 1/2$ for all j when n is large enough. By working with such n exclusively, we may and do assume that $x_{nj} \leq 1/2$ for all n and j. For x in $[0, 1/2]$, Taylor's theorem shows that

$$|x + \log(1 - x)| = \sum_2^\infty \frac{x^m}{m} \leq \frac{1}{2}x^2 \frac{1}{1 - x} \leq x^2\ .$$

Thus,

$$|\sum_j x_{nj} + \sum_j \log(1 - x_{nj})| \leq \sum_j x_{nj}^2 \leq (\max_j x_{nj}) \sum_i x_{ni} \to 0\ .$$

Since $\lim \sum_j x_{nj} = b = -\log e^{-b}$, this shows that

$$\lim_n |\log \prod_i (1 - x_{nj}) - \log e^{-b}| = 0 \ ,$$

which is equivalent to the desired conclusion. □

The preceding is very close to Lindeberg's theorem, at least in spirit. Take n very large. The probability that $X_{nj} \geq 2$ for some j is basically zero by the assumption that $\sum_j \mathbb{P}\{X_{nj} \geq 2\} \to 0$. Thus, the contribution of X_{nj} to the sum Z_n is either 0 or 1, that is, the X_{nj} are nearly Bernoulli variables. Moreover, the probability that the contribution is 1 is very small since $\mathbb{P}\{X_{nj} \geq 1\} \to 0$. Thus, the sum Z_n consists of a large number of nearly Bernoulli variables whose success probabilities are uniformly small, but the sum of those probabilities is $\mathbb{E}\ Z_n$, which is roughly c. So, the situation resembles that for the convergence of the binomial distribution to the Poisson distribution; see Exercise 5.28.

Convergence to infinitely divisible variables

A random variable Z is said to be *infinitely divisible* provided that, for each integer $n \geq 1$, there are independent and identically distributed random variables $Z_{n,1}, \ldots, Z_{n,n}$ whose sum has the same distribution as Z. A probability measure is said to be *infinitely divisible* if it is the distribution of an infinitely divisible variable. It follows that the Fourier transform f of such a measure is characterized by the property that $f^{1/n}$ is the Fourier transform of a probability measure for each n.

Poisson distributions, Gaussian distributions, and gamma distributions are all infinitely divisible; see Exercise II.5.22 for a proof. It is obvious that the sums of infinitely divisible random variables are infinitely divisible. We shall show later that, in fact, every infinitely divisible random variable is the sum of a Gaussian variable and a limit of linear combinations of independent Poisson variables.

The theorems above have shown that the row sums of the array 8.2 converge in distribution to Gaussian or Poisson variables under certain conditions. The following shows that, in general, the limits are infinitely divisible.

8.26 THEOREM. *Suppose that the array 8.2 is strictly triangular (that is, $k_n = n$) and that $X_{n1}, \ldots, X_{nn}$ are independent and identically distributed. If (Z_n) converges in distribution to some variable Z, then Z is infinitely divisible. Conversely, if Z is an infinitely divisible random variable, then there is an array $[X_{ni}]$ whose row sums Z_n converge in distribution to Z.*

Proof. The converse statement is trivial: if Z is infinitely divisible, there are independent and identically distributed variables $X_{nj}, j = 1, \ldots, n$, whose sum has the same distribution as Z; then, we define the array by these X_{nj} and the row sums Z_n have the same distribution as Z for all n.

Assume that the array is as described and that $Z_n \to Z$ in distribution; we need to show that Z is infinitely divisible. Fix m; we need to show that there are independent and identically distributed variables $Y_1, \ldots, Y_m$ whose sum has the same distribution as Z.

Let $n_k = mk$, $k \geq 1$. Consider the n_k^{th} row of the array and let $S_{1,k}$ be the sum of the first k variables there, $S_{2,k}$ the sum of the next k variables, and so on, ending with $S_{m,k}$, the sum of the last k variables at or before the diagonal. Then, these m sums are independent and identically distributed and their sum is $Z_{n_k} = Z_{mk}$.

In Lemma 8.27 below we shall show that there is a subsequence K such that $(S_{1,k})$ converges in distribution along K to some random variable Y_1. Then, since S_{jk} has the same distribution as $S_{1,k}$, the sequence (S_{jk}) converges in distribution along the same K to some variable Y_j for $j = 1, \ldots, m$, and moreover the variables $Y_1, \ldots, Y_m$ are identically distributed and can be made to be independent. The independence of the $S_{j,k}$ for $j = 1, \ldots, m$ implies (see Exercise 5.22) that

$$Z_{mk} = S_{1,k} + \cdots + S_{m,k} \to Y_1 + \cdots + Y_m$$

in distribution as $k \to \infty$ along K. In other words, (Z_n) has a subsequence that converges to $Y_1 + \cdots + Y_m$ in distribution. Since $Z_n \to Z$ in distribution by assumption, this implies that Z has the same distribution as $Y_1 + \cdots + Y_m$, and the proof is complete. □

The Lemma needed in the proof is next, with the same assumptions and notations.

8.27 LEMMA. *The sequence $(S_{1,k})$ has a subsequence that is convergent in distribution.*

Proof. In view of Theorem 5.13 on tightness and existence of subsequences that converge, it is enough to show that the distributions of $S_{1,k}$, $k \geq 1$, form a tight family.

Suppose, to the contrary, that the family is not tight. Then, there exists $\varepsilon > 0$ such that for every b in $\mathbb{R}_+$ there is a subsequence J such that $\mathbb{P}\{S_{1,k} \in [-b, b]\} \leq 1 - 2\varepsilon$ for every k in J, which implies that either $\mathbb{P}\{S_{1,k} > b\}$ or $\mathbb{P}\{S_{1,k} < -b\}$ exceeds ε. Since $S_{1,k}, \ldots, S_{m,k}$ are independent and identically distributed and sum to Z_{mk}, it follows that either $\mathbb{P}\{Z_{mk} > mb\}$ or $\mathbb{P}\{Z_{mk} < -mb\}$ exceeds ε^m. Thus,

$$\mathbb{P}\{|Z_{mk}| > mb\} \geq \varepsilon^m$$

for every k in J. Since (Z_n) converges in distribution to Z, this implies that $\mathbb{P}\{|Z| > mb\} \geq \varepsilon^m$ for every b such that mb and $-mb$ are points of continuity for the distribution function of Z. Letting $b \to \infty$ we see that $\mathbb{P}\{S \in \mathbb{R}\} \leq 1 - \varepsilon^m < 1$, which contradicts the definition of convergence in distribution. □

Finally, returning to the array 8.2 with the assumptions stated there, we state the following general theorem which also identifies the limiting distribution. We do not assume identical distribution for the elements of a row. Recall that ∂A denotes the boundary of A. We omit the proof.

8.28 THEOREM. *Suppose that the array $[X_{nj}]$ satisfies the "infinitesimality" condition 8.22. In order that $(Z_n - a_n)$ converge in distribution for some (a_n) in $\mathbb{R}$, it is necessary and sufficient that there exist a constant $b \geq 0$ and a measure λ on the Borel σ-algebra of $\mathbb{R}$ satisfying*

8.29 $$\int_{\mathbb{R}} \lambda(dx)(x^2 \wedge 1) < \infty$$

such that

$$\lim_n \sum_j \mathbb{P}\{X_{nj} \in A\} = \lambda(A)$$

for every Borel set A with $\lambda(\partial A) = 0$, and

$$\lim_{\varepsilon \to 0} \liminf_{n\to\infty} \sum_j \text{Var } X_{nj} \bar{i}_\varepsilon \circ |X_{nj}| = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sum_j \text{Var } X_{nj} \bar{i}_\varepsilon \circ |X_{nj}| = b \ .$$

In the case of convergence, the constants a_n can be chosen as

$$a_n = \sum_j \mathbb{E}\ X_{nj} 1_{[-c,c]} \circ X_{nj}$$

where $c > 0$ is an arbitrary constant such that neither c nor $-c$ is an atom of λ. With this choice of (a_n) and assuming convergence, the limit variable Z is infinitely divisible and its characteristic function is

$$\exp\left\{ -\frac{1}{2} r^2 b + \int_{\mathbb{R}} \lambda(dx) \left[e^{irx} - 1 - irx 1_{[-c,c]}(x) \right] \right\} , \quad r \in \mathbb{R} \ .$$

In Chapter VII we shall construct infinitely divisible variables. The measure λ appearing in the theorem is called the Lévy measure; it is defined to satisfy 8.29, which is used to show that certain series converge.

Exercises

8.30 *Lyapunov's condition.* Suppose that

$$\lim_n \sum_j \mathbb{E}|X_{nj}|^{2+\delta} = 0$$

for some $\delta > 0$. Show that this implies Lindeberg's condition.

8.31 *Lindeberg's theorem.* The classical case of Theorem 8.1 can be put in the form of an array 8.2 by defining $X_{nj} = X_j/\sqrt{n}$, $j = 1, \ldots, n$, for some sequence (X_n) of independent and identically distributed variables. Show that, assuming Var $X_n = b < \infty$, then, Lindeberg's condition holds. So, Lindeberg's theorem subsumes the classical case.

8.32 *Convergence to Poisson.* Suppose that the nth row of the array 8.2 is $(X_{n,1},\ \ldots,\ X_{n,n},\ 0, 0,\ \ldots)$ where the X_{nj} are independent and identically distributed $\mathbb{N}$-valued variables with

$$p_n = \mathbb{P}\{X_{nj} \geq 1\}\ , \quad q_n = \mathbb{P}\{X_{nj} \geq 2\}\ .$$

If $nq_n \to 0$ and $np_n \to c$ for some constant $c > 0$, then $Z_n \to \mathfrak{P}_c$ in distribution. Show this directly.

Chapter IV

Conditioning

This chapter is a continuation of Chapter II. We start with conditional expectations, namely, estimation of a random variable in the presence of partial information. Then, the notion of conditional independence follows as a natural generalization of independence, where the role played by the expectation operator is now played by the conditional expectation operator. Finally, we discuss various ways of constructing random variables and stochastic processes.

1 Conditional Expectations

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$, and X an $\bar{\mathbb{R}}$-valued random variable. As usual, we regard $\mathcal{F}$ both as a collection of events and as the collection of all $\mathcal{F}$-measurable random variables. Moreover, we think heuristically and regard $\mathcal{F}$ as a body of information, namely, the information that determines the values $V(\omega)$ for all V in $\mathcal{F}$ and all ω in Ω. Recall from Chapter II, Section 4, that knowing the value $V(\omega)$ is much less than knowing ω but, rather, it affords us to infer various properties of ω.

Heuristically, the conditional expectation of X given $\mathcal{F}$ is a random variable $\bar{X}$ such that, for every possibility ω, $\bar{X}(\omega)$ is the "best" estimate of $X(\omega)$ based on the information $\mathcal{F}$. So, $\bar{X}$ must be determined by $\mathcal{F}$, and among all random variables determined by $\mathcal{F}$ it must be the "closest" to X, the terms "best" and "closest" having a definite meaning at least when X is square integrable, that is, when $\mathbb{E}X^2 < \infty$. Then, $\mathbb{E}(X - Y)^2$ can be regarded as a measure of the error committed by using Y as an estimate of X, and we want $\mathbb{E}(X - \bar{X})^2 \le \mathbb{E}(X - Y)^2$ for all Y in $\mathcal{F}$.

Preparatory steps

These are to illustrate the thought processes involved and to motivate the formal definitions to come. Recall that $\mathcal{H}$ is the collection of all events

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_4,

and is also the collection of all $\bar{\mathbb{R}}$-valued random variables, $\mathcal{H}_+$ being the subcollection consisting of the positive ($\bar{\mathbb{R}}_+$-valued) variables. Similarly, for the sub-σ-algebra $\mathcal{F}$, we regard $\mathcal{F}$ as the collection of all $\mathcal{F}$-measurable random variables and $\mathcal{F}_+$ as the sub-collection of positive ones. To simplify the discussion here, we assume $X \in \mathcal{H}_+$.

Let H be an event. Fix ω in Ω. Suppose that all that is known about ω is that it is in H. Based on this information, our best estimate of $X(\omega)$ should be the "average" of X over H, namely, the number

1.1 $$\mathbb{E}_H X = \frac{1}{\mathbb{P}(H)} \int_H X \, d\mathbb{P} = \frac{1}{\mathbb{P}(H)} \mathbb{E} X 1_H$$

assuming that $\mathbb{P}(H) > 0$. If $\mathbb{P}(H) = 0$ then so is the integral over H and we allow $\mathbb{E}_H X$ to be any positive number desired. This estimate is best in the same sense that $\mathbb{E}X$ is the best number estimating $X(\omega)$ when nothing whatsoever is known about ω, which corresponds to the case $H = \Omega$ (see Exercise II.2.15). The number $\mathbb{E}_H X$ is called the *conditional expectation* of *X given the event H.*

Suppose, next, that $\mathcal{F}$ is generated by a measurable partition (H_n) of Ω. For fixed ω again, consider what our estimate of $X(\omega)$ should be if we were given the information $\mathcal{F}$. Given $\mathcal{F}$, we shall be able to tell which one of the events $H_1, H_2, \ldots$ includes ω, and if it were H_n that included ω then our estimate $\bar{X}(\omega)$ would be $\mathbb{E}_{H_n} X$. In other words,

1.2 $$\bar{X}(\omega) = \sum_n (\mathbb{E}_{H_n} X) 1_{H_n}(\omega).$$

Doing this thinking for each possibility ω, we arrive at a random variable $\bar{X}$, which we denote by $\mathbb{E}_{\mathcal{F}} X$ and call the *conditional expectation of X given $\mathcal{F}$.* Figure 3 is for the case where $\Omega = (0, 1)$, $\mathcal{H}$ is the Borel σ-algebra, and $\mathbb{P}$ is the Lebesgue measure on Ω, and $\mathcal{F} = \sigma\{H_1, H_2, H_3\}$.

To see the proper generalization for arbitrary $\mathcal{F}$ we mark two properties of $\bar{X} = \mathbb{E}_{\mathcal{F}} X$ for the special $\mathcal{F}$ above. First, $\bar{X}$ belongs to $\mathcal{F}$, that is, $\bar{X}$ is determined by the information $\mathcal{F}$. Second, $\mathbb{E}\, VX = \mathbb{E}\, V\bar{X}$ for every V in $\mathcal{F}_+$; this is obvious from 1.1 and 1.2 when $V = 1_{H_n}$ for some fixed n, and it extends to arbitrary V in $\mathcal{F}_+$ through the monotone convergence theorem, because every such V has the form $V = \sum a_n 1_{H_n}$ for this $\mathcal{F}$. We use these two properties to define conditional expectations and proceed to show their existence, uniqueness, and various properties.

Definition of conditional expectations

1.3 DEFINITION. *Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$. The* conditional expectation of X given $\mathcal{F}$, *denoted by $\mathbb{E}_{\mathcal{F}} X$, is defined in two steps: For X in $\mathcal{H}_+$, it is defined to be any random variable $\bar{X}$ that satisfies*

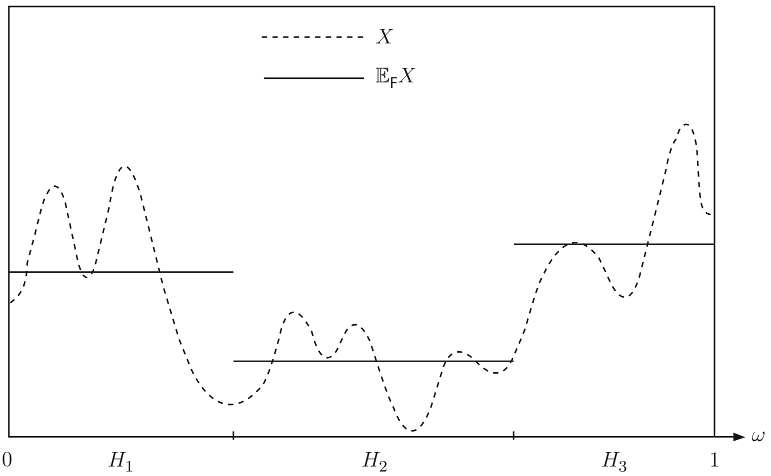


Figure 3: The conditional expectation $\mathbb{E}_{\mathcal{F}}X$ is that $\mathcal{F}$-measurable random variable that is closest to X.

1.4 a) *measurability:* $\bar{X}$ belongs to $\mathcal{F}_+$, and
b) *projection property:* $\mathbb{E}\ VX = \mathbb{E}\ V\bar{X}$ for every V in $\mathcal{F}_+$,

and then we write $\mathbb{E}_{\mathcal{F}}X = \bar{X}$ *and call* $\bar{X}$ *a version of* $\mathbb{E}_{\mathcal{F}}X$. *For arbitrary* X *in* $\mathcal{H}$, *if* $\mathbb{E}X$ *exists, then we define*

1.5 $$\mathbb{E}_{\mathcal{F}}X = \mathbb{E}_{\mathcal{F}}X^+ - \mathbb{E}_{\mathcal{F}}X^-;$$

otherwise, if $\mathbb{E}X^+ = \mathbb{E}X^- = +\infty$, *then* $\mathbb{E}_{\mathcal{F}}X$ *is left undefined.*

1.6 Remarks. a) *Projection property.* For X in $\mathcal{H}_+$, the projection property 1.4b for $\bar{X} = \mathbb{E}_{\mathcal{F}}X$ is equivalent to the condition that

$$\mathbb{E}\ 1_H X = \mathbb{E}\ 1_H \bar{X}, \qquad H \in \mathcal{F}.$$

This is immediate from the monotone convergence theorem used on both sides of the equality to extend it from indicators to simple variables and from simple to arbitrary positive V in $\mathcal{F}$.

b) *Tests for equality.* This is to recall an elementary point made in Remark II.2.3d. If Y and Z are in $\mathcal{F}_+$ and if

$$\mathbb{E}\ VY \le \mathbb{E}\ VZ \qquad \text{for every } V \text{ in } \mathcal{F}_+$$

or for every indicator 1_H in $\mathcal{F}_+$, then $Y \le Z$ almost surely (and the almost sure set $\{Y \le Z\}$ is in $\mathcal{F}$). If the inequality is in fact an equality for all V in $\mathcal{F}_+$, then $Y = Z$ almost surely.

c) *Uniqueness of conditional expectations.* Let $\bar{X}$ and $\bar{\bar{X}}$ be versions of $\mathbb{E}_{\mathcal{F}}X$ for $X \geq 0$. Then they are both in $\mathcal{F}_+$ and $\mathbb{E}\ V\bar{X} = \mathbb{E}\ VX = \mathbb{E}\ V\bar{\bar{X}}$ for every V in $\mathcal{F}_+$. It follows from the preceding remark that $\bar{X} = \bar{\bar{X}}$ almost surely. Conversely, if $\mathbb{E}_{\mathcal{F}}X = \bar{X}$, and if $\bar{\bar{X}} \in \mathcal{F}_+$ and $\bar{\bar{X}} = \bar{X}$ almost surely, then $\bar{\bar{X}}$ satisfies 1.4 and hence is another version of $\mathbb{E}_{\mathcal{F}}X$. This uniqueness up to equivalence extends to $\mathbb{E}_{\mathcal{F}}X$ for arbitrary X for which $\mathbb{E}X$ exists; see also (f) below.

d) *Language.* In view of the uniqueness up to equivalence, the definite article in "the conditional expectation ..." is only a slight abuse of language. For the same reason, $\mathbb{E}_{\mathcal{F}}X$ should be regarded as an all-purpose notation for each and every version $\bar{X}$. Some authors take the extra logical step and define "the conditional expectation ..." to be the equivalence class of all versions, and then use $\mathbb{E}_{\mathcal{F}}X$ as a representative of that class, and write "$\mathbb{E}_{\mathcal{F}}X = \bar{X}$ almost surely" to mean that $\bar{X}$ is a version.

e) *Integrability.* If $X \in \mathcal{H}_+$ then $\mathbb{E}\ X = \mathbb{E}\ \mathbb{E}_{\mathcal{F}}X$ in view of the projection property with $V = 1$; thus, if X is integrable then so is $\mathbb{E}_{\mathcal{F}}X$. This remark applies to X^+ and X^- separately for arbitrary X and, thus, if X is integrable then so is $\mathbb{E}_{\mathcal{F}}X$. Hence, if X is integrable and $\mathbb{E}_{\mathcal{F}}X = \bar{X}$, the projection property can be expressed as

$$\mathbb{E}\ V(X - \bar{X}) = 0 \qquad \text{for every bounded } V \text{ in } \mathcal{F}.$$

In general, if X is not integrable, this expression fails to have meaning because the expectation involved might fail to exist.

f) *Definition for arbitrary X.* If X^+ is integrable, then so is $\mathbb{E}_{\mathcal{F}}X^+$ which implies that $\mathbb{E}_{\mathcal{F}}X^+$ is real-valued almost surely, and a similar statement holds for X^-. Thus, if $\mathbb{E}X$ exists, there is an almost sure event Ω_0 in $\mathcal{F}$ such that, on Ω_0, at least one of the random variables $\mathbb{E}_{\mathcal{F}}X^+$ and $\mathbb{E}_{\mathcal{F}}X^-$ is real-valued, and therefore, $\bar{X} = \mathbb{E}_{\mathcal{F}}X^+ - \mathbb{E}_{\mathcal{F}}X^-$ is well-defined on Ω_0. By putting $\bar{X}$ equal to 0 outside Ω_0 for definiteness, we obtain a well-defined version $\bar{X}$ of $\mathbb{E}_{\mathcal{F}}X$.

g) *Heuristics.* Suppose that X is integrable; then, so is $\bar{X} = \mathbb{E}_{\mathcal{F}}X$ and so is $\tilde{X} = X - \bar{X}$. We thus have a decomposition

$$X = \bar{X} + \tilde{X}$$

where $\bar{X}$ is determined by the information $\mathcal{F}$, and $\tilde{X}$ is orthogonal to $\mathcal{F}$ (that is, $\mathbb{E}\ 1_H\tilde{X} = 0$ for all events H in $\mathcal{F}$; see the remark (e) above). For this reason, we may call $\bar{X}$ the orthogonal projection of X onto $\mathcal{F}$, and the defining property 1.4b is named "projection property" to suggest this picture. We shall return to this theme and make it rigorous in Theorem 1.11.

Existence of conditional expectations

The following proposition uses the Radon-Nikodym theorem to show the existence of conditional expectations. We limit the proposition to positive X, since the extension to arbitrary X is covered by Remark 1.6f above. See 1.15 below for another proof which is independent of the Radon-Nikodym theorem.

1.7 THEOREM. *Let $X \in \mathcal{H}_+$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$. Then $\mathbb{E}_{\mathcal{F}}X$ exists and is unique up to equivalence.*

Proof. For each event H in $\mathcal{F}$, define

$$P(H) = \mathbb{P}(H), \qquad Q(H) = \int_H \mathbb{P}(d\omega)X(\omega).$$

On the measurable space $(\Omega, \mathcal{F})$, then, P is a probability measure, and Q is a measure that is absolutely continuous with respect to P. Hence, by I.5.11, the Radon-Nikodym theorem, there exists $\bar{X}$ in $\mathcal{F}_+$ such that

$$\int_\Omega Q(d\omega)V(\omega) = \int_\Omega \mathbb{P}(d\omega)\bar{X}(\omega)V(\omega)$$

for every V in $\mathcal{F}_+$. This shows that $\bar{X}$ is a version of $\mathbb{E}_{\mathcal{F}}X$. For its uniqueness up to almost sure equality, see Remark 1.6c. □

Properties similar to expectations

The following properties are the same as those for expectations. They are easy to show and will not be proved. Throughout, $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{H}$, all the random variables are in $\mathcal{H}$ (that is, they are arbitrary $\bar{\mathbb{R}}$-valued variables), the constants a, b, c are real numbers, *and* it is assumed that all conditional expectations involved exist. Of course, all these conditional expectations exist if all the random variables are positive or integrable.

1.8 *Monotonicity:* $\quad X \le Y \;\Rightarrow\; \mathbb{E}_{\mathcal{F}}X \le \mathbb{E}_{\mathcal{F}}Y.$

Linearity: $\quad \mathbb{E}_{\mathcal{F}}(aX + bY + c) = a\,\mathbb{E}_{\mathcal{F}}X + b\,\mathbb{E}_{\mathcal{F}}Y + c.$

Monotone convergence theorem: $\quad X_n \ge 0, X_n \nearrow X \;\Rightarrow\; \mathbb{E}_{\mathcal{F}}X_n \nearrow \mathbb{E}_{\mathcal{F}}X.$

Fatou's lemma: $\quad X_n \ge 0 \;\Rightarrow\; \mathbb{E}_{\mathcal{F}} \liminf X_n \;\le\; \liminf \mathbb{E}_{\mathcal{F}}X_n.$

Dominated convergence theorem: $\quad X_n \to X\;,\; |X_n| \le Y,\; Y$ integrable $\;\Rightarrow\; \mathbb{E}_{\mathcal{F}}X_n \to \mathbb{E}_{\mathcal{F}}X.$

Jensen's inequality: $\quad f$ convex $\;\Rightarrow\; \mathbb{E}_{\mathcal{F}}f(X) \ge f(\mathbb{E}_{\mathcal{F}}X).$

1.9 REMARK. There are occasions when the properties above require careful interpretation, because conditional expectations are unique only up to equivalence. To illustrate, we now re-state the monotone convergence theorem for them in precise terms and prove it:

> Suppose that $(X_n) \subset \mathcal{H}_+$ and $X_n \nearrow X$. Then there are versions $\bar{X}_n$ of $\mathbb{E}_{\mathcal{F}}X_n$ and $\bar{X}$ of $\mathbb{E}_{\mathcal{F}}X$ such that $\bar{X}_n \nearrow \bar{X}$.

To show this, let X'_n be a version of $\mathbb{E}_{\mathcal{F}}X_n$ for each n. By monotonicity, for each n, $X'_n \le X'_{n+1}$ almost surely. Thus, there is an almost sure event

Ω_0 in $\mathcal{F}$ such that, for every ω in Ω_0, we have $X_1'(\omega) \le X_2'(\omega) \le \cdots$. Define $\bar{X}_n(\omega) = X_n'(\omega)$ for ω in Ω_0 and put $\bar{X}_n(\omega) = 0$ for ω outside Ω_0. Then, each $\bar{X}_n$ is a version of $\mathbb{E}_{\mathcal{F}} X_n$, and $(\bar{X}_n)$ is an increasing sequence in $\mathcal{F}_+$; let $\bar{X}$ be the limit. Then, $\bar{X} \in \mathcal{F}_+$ and, for $V \in \mathcal{F}_+$,

$$\mathbb{E}\ V\bar{X} = \lim \mathbb{E}\ V\bar{X}_n = \lim \mathbb{E}\ VX_n = \mathbb{E}\ VX$$

where the projection property 1.4b justifies the middle equality sign, and the ordinary monotone convergence theorem the other two. Thus, $\bar{X}$ is a version of $\mathbb{E}_{\mathcal{F}} X$.

Special properties

The following theorem summarizes the properties special to conditional expectations. Heuristically, conditional determinism is that, if W is determined by $\mathcal{F}$ then it should be treated as if it is a deterministic number. For the repeated conditioning, think of $\mathcal{F}$ as the information a fool has, and $\mathcal{G}$ as that a genius has: the genius cannot improve on the fool's estimate, but the fool has no difficulty worsening the genius's. In repeated conditioning, fools win all the time.

1.10 THEOREM. *Let $\mathcal{F}$ and $\mathcal{G}$ be sub-σ-algebras of $\mathcal{H}$. Let W and X be random variables (in $\mathcal{H}$) such that $\mathbb{E}X$ and $\mathbb{E}WX$ exist. Then, the following hold:*

a) Conditional determinism*:* $W \in \mathcal{F} \;\Rightarrow\; \mathbb{E}_{\mathcal{F}} WX = W\mathbb{E}_{\mathcal{F}} X$.
b) Repeated conditioning*:* $\mathcal{F} \subset \mathcal{G} \;\Rightarrow\; \mathbb{E}_{\mathcal{F}}\mathbb{E}_{\mathcal{G}} X = \mathbb{E}_{\mathcal{G}}\mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}} X$.

Proof. We give the proofs for W and X positive; the general cases follow by easy considerations.

a) Suppose $X \in \mathcal{H}_+$ and $W \in \mathcal{F}_+$. Then $\bar{X} = \mathbb{E}_{\mathcal{F}} X$ is in $\mathcal{F}_+$ and

$$\mathbb{E}\ V \cdot (WX) = \mathbb{E}\ (V \cdot W)X = \mathbb{E}\ (V \cdot W)\bar{X} = \mathbb{E}\ V \cdot (W\bar{X})$$

for every V in $\mathcal{F}_+$, where the crucial middle equality sign is justified by the projection property for $\bar{X}$ and the fact that $VW \in \mathcal{F}_+$. Hence, $W\bar{X} = W\mathbb{E}_{\mathcal{F}} X$ is a version of $\mathbb{E}_{\mathcal{F}}(WX)$.

b) Let $\mathcal{F} \subset \mathcal{G}$ and $X \in \mathcal{H}_+$. Since $W = \mathbb{E}_{\mathcal{F}} X$ is in $\mathcal{F}_+$ by definition, we have that $W \in \mathcal{G}_+$ and hence $\mathbb{E}_{\mathcal{G}} W = W$ by the conditional determinism property proved above. This proves the second equality in the statement of repeated conditioning. There remains to show that, with $Y = \mathbb{E}_{\mathcal{G}} X$ and $\bar{X} = \mathbb{E}_{\mathcal{F}} X$,

$$\mathbb{E}_{\mathcal{F}} Y = \bar{X}.$$

Obviously, $\bar{X} \in \mathcal{F}_+$. To check the projection property, let $V \in \mathcal{F}_+$. By the definition of $\bar{X}$, we have $\mathbb{E}\ V\bar{X} = \mathbb{E}\ VX$. By the definition of Y, we have $\mathbb{E}\ VY = \mathbb{E}\ VX$, since $V \in \mathcal{G}_+$, which is in turn because $V \in \mathcal{F}_+$ and $\mathcal{F} \subset \mathcal{G}$. Hence, $\mathbb{E}\ V\bar{X} = \mathbb{E}\ VY$ as needed. □

Conditioning as projection

We return to the beginnings. Recall the heuristic remarks about $\mathbb{E}_{\mathcal{F}}X$, its geometric interpretation as a projection, and its interpretation as an $\mathcal{F}$-determined estimate that minimizes the expected value of the squared error caused by such estimates of X. The following theorem is the rigorous statement justifying such remarks. As a corollary, we obtain a second proof of the existence and uniqueness Theorem 1.7, this time without recourse to the unproved Radon-Nikodym theorem.

Since we shall be dealing with "expected value of squared error," we are necessarily limited to square-integrable random variables in the next theorem. Accordingly, if $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{H}$, we write $L^2(\mathcal{F})$ for the collection of all square-integrable random variables in $\mathcal{F}$; then, $L^2(\mathcal{H})$ is the L^2-space introduced in Chapter II.

1.11 THEOREM. *For every X in $L^2(\mathcal{H})$ there exists a unique (up to equivalence) $\bar{X}$ in $L^2(\mathcal{F})$ such that*

1.12 $$\mathbb{E}\,|X-\bar{X}|^2 = \inf_{Y\in L^2(\mathcal{F})}\mathbb{E}\,|X-Y|^2$$

Moreover, $X-\bar{X}$ is orthogonal to $L^2(\mathcal{F})$, that is, for every V in $L^2(\mathcal{F})$,

1.13 $$\mathbb{E}\,V\cdot(X-\bar{X}) = 0$$

1.14 REMARKS. Obviously, $\bar{X} = \mathbb{E}_{\mathcal{F}}X$. With $\mathbb{E}XY$ as the inner product of X and Y, the space $L^2(\mathcal{H})$ is a complete Hilbert space, and $L^2(\mathcal{F})$ is a subspace of it. In the terminology of such spaces, $\bar{X}$ is the orthogonal projection of the vector X onto $L^2(\mathcal{F})$, and we have the decomposition $X = \bar{X}+\tilde{X}$ where $\bar{X}$ is in $L^2(\mathcal{F})$ and $\tilde{X}$ is orthogonal to $L^2(\mathcal{F})$.

Proof. It is convenient to use the L^2-norm introduced in Chapter II, section 3, but omitting the subscripts: Thus, $\|X\| = \|X\|_2 = \sqrt{\mathbb{E}X^2}$. Fix X in $L^2(\mathcal{H})$. Define

$$\delta = \inf_{Y\in L^2(\mathcal{F})}\|X-Y\|$$

and let $(Y_n)\subset L^2(\mathcal{F})$ such that $\delta_n = \|X-Y_n\| \to \delta$ as $n\to\infty$.

We show now that (Y_n) is Cauchy for convergence in $L^2(\mathcal{F})$. Note that

$$|Y_n-Y_m|^2 = 2\,|X-Y_m|^2 + 2\,|X-Y_n|^2 - 4\,|X-\tfrac{1}{2}(Y_m+Y_n)|^2$$

and take expectations to get

$$\mathbb{E}|Y_n-Y_m|^2 \le 2\delta_m^2 + 2\delta_n^2 - 4\delta^2,$$

since $\|X-Y\|\ge\delta$ for every Y in $L^2(\mathcal{F})$ and in particular for $Y=\frac{1}{2}(Y_m+Y_n)$.

Hence, (Y_n) is Cauchy, and by Theorem III.4.6 and III.4.13, there exists $\bar{X}$ in $L^2(\mathcal{F})$ such that $\|Y_n-\bar{X}\|\to 0$ as $n\to\infty$. Of course, $\bar{X}$ is unique up

to almost sure equality. Since $\bar{X} \in L^2(\mathcal{F})$, we have $\|X - \bar{X}\| \geq \delta$; and by Minkowski's inequality (see Theorem II.3.6),

$$\|X - \bar{X}\| \leq \|X - Y_n\| + \|Y_n - \bar{X}\| \longrightarrow \delta + 0 = \delta.$$

It follows that $\|X - \bar{X}\| = \delta$ as needed to complete the proof of the first statement.

Moreover, for V in $L^2(\mathcal{F})$ and a real, since $\mathbb{E}(X - \bar{X})^2 = \delta^2$,

$$a^2\,\mathbb{E}\,V^2 - 2a\,\mathbb{E}\,V\cdot(X - \bar{X}) + \delta^2 = \|aV - (X - \bar{X})\|^2 = \|X - (aV + \bar{X})\|^2 \geq \delta^2$$

because $aV + \bar{X} \in L^2(\mathcal{F})$. Thus, $a^2\,\mathbb{E}\,V^2 - 2a\,\mathbb{E}\,V\cdot(X - \bar{X}) \geq 0$ for every real number a, which is impossible unless 1.13 holds. □

1.15 SECOND PROOF FOR 1.7. Let $X \in \mathcal{H}_+$. Define $X_n = X \wedge n$, which is in $L^2(\mathcal{H})$. Thus, by the preceding theorem, there is $\bar{X}_n$ in $L^2(\mathcal{F})$ such that, for every V in $L^2(\mathcal{F})$,

1.16 $$\mathbb{E}\,VX_n = \mathbb{E}\,V\bar{X}_n$$

Now fix $V = 1_H$ for some event H in $\mathcal{H}$. Then, since $X_n \nearrow X$, 1.16 implies that $(\bar{X}_n)$ increases almost surely to some $\bar{X}$ in $\mathcal{F}_+$, and the monotone convergence theorem shows that $\mathbb{E}VX = \mathbb{E}V\bar{X}$. □

Conditional expectations given random variables

Let Y be a random variable taking values in some measurable space. Recall that σY denotes the σ-algebra generated by Y, which consists of numerical random variables of the form $f \circ Y$ for some measurable function f; see II.4.1 *et seq.* For X in $\mathcal{H}$, the *conditional expectation of X given Y* is defined to be $\mathbb{E}_{\sigma Y} X$. Similarly, if $\{Y_t : t \in T\}$ is a collection of random variables taking values in some measurable spaces, the *conditional expectation of X given* $\{Y_t : t \in T\}$ is $\mathbb{E}_{\mathcal{F}} X$ with $\mathcal{F} = \sigma\{Y_t : t \in T\}$. Of course, these two definitions coincide when the collection $\{Y_t : t \in T\}$ is identified with $Y = (Y_t)_{t \in T}$. The following is an immediate consequence of the definitions here and Theorem II.4.4.

1.17 THEOREM. *Let $X \in \mathcal{H}_+$. Let Y be a random variable taking values in some measurable space $(E, \mathcal{E})$. Then, every version of $\mathbb{E}_{\sigma Y} X$ has the form $f \circ Y$ for some f in $\mathcal{E}_+$. Conversely, $f \circ Y$ is a version of $\mathbb{E}_{\sigma Y} X$ if and only if*

1.18 $$\mathbb{E}\,f\circ Y\;h\circ Y = \mathbb{E}\,X \cdot h\circ Y \qquad \textit{for every } h \textit{ in } \mathcal{E}_+.$$

In the preceding theorem, the hypothesis that X be positive is inessential, except for ensuring that the conditional expectation exists. If X is integrable, the claim remains the same for f in $\mathcal{E}$ and bounded h in $\mathcal{E}_+$. Also, when X is integrable, 1.18 can be replaced, via a monotone class – monotone convergence argument, with the requirement that it hold for $h = 1_E$ and $h = 1_A$ for every A in some p-system $\mathcal{E}_0$ that generates $\mathcal{E}$.

Notation

The traditional notation for $\mathbb{E}_{\mathcal{F}}X$ is $\mathbb{E}(X|\mathcal{F})$. The notation $\mathbb{E}_{\mathcal{F}}X$ is sharper, conveys better its meaning as a projection, is clearer in expressions like $\mathbb{E}_{\mathcal{F}}\mathbb{E}_{\mathcal{G}}$, and conforms to ordinary practices in notations for operators. The other, $\mathbb{E}(X|\mathcal{F})$ is no better than a shorthand, but has the advantage of being more linear and more convenient when $\mathcal{F}$ has to be replaced by a cumbersome expression. For instance, it is usual to write $\mathbb{E}(X|Y_t : t \in T)$ for $\mathbb{E}_{\mathcal{F}}X$ when $\mathcal{F} = \sigma\{Y_t : t \in T\}$.

The notations $\mathbb{E}(X|Y)$ and $\mathbb{E}_{\sigma Y}X$ are used often to denote $\mathbb{E}_{\mathcal{F}}X$ with $\mathcal{F} = \sigma Y$. Similarly, $\mathbb{E}(X|H)$ is the traditional notation for $\mathbb{E}_H X$ and works better when H is expressed in terms of other things, for example, if $H = \{Y = y\}$.

Finally, the notation $\mathbb{E}(X|Y = y)$ is used and read "the conditional expectation of X given that $Y = y$" despite its annoying ambiguity. It is reasonable, when $\mathbb{P}\{Y = y\} > 0$, as a notation for $\mathbb{E}_H X = \mathbb{E}(X|H)$ with $H = \{Y = y\}$. It is also used when $\mathbb{P}\{Y = y\} = 0$ for all y, and the proper interpretation then is that it is a notation for $f(y)$ when $f \circ Y = \mathbb{E}_{\sigma Y}X$.

Examples

1.19 *Remaining lifetime.* A device is installed at time 0. It is known that the device has not failed during $[\,0\,,\ t\,]$. We would like to estimate its remaining lifetime. Let X represent the length of its life. What we know is that the event $H = \{X > t\}$ has occurred (that is, outcome is a point in H). We want to compute $\mathbb{E}_H(X - t) = \mathbb{E}_H X - t$:

Let μ be the distribution of X. Then, with $H = \{X > t\}$,

$$\mathbb{E}_H X = \frac{1}{\mathbb{P}(H)}\mathbb{E}\; X 1_H = \frac{1}{\mu(t,\infty)}\int_{\mathbb{R}_+}\mu(ds)\; s\; 1_{(t,\infty)}(s) = \frac{1}{\mu(t,\infty)}\int_{(t,\infty)}\mu(ds)s.$$

1.20 *Effect of independence.* Suppose that X and Y are independent and take values in $(E, \mathcal{E})$ and $(D, \mathcal{D})$ respectively. If $f \in \mathcal{E}_+$ and $g \in \mathcal{D}_+$, then

1.21 $$\mathbb{E}_{\sigma Y}\; f\circ X\; g\circ Y = g\circ Y\; \mathbb{E}\; f\circ X$$

This adds meaning to independence: information obtained by observing Y determines $g \circ Y$, but is worthless for estimating $f \circ X$. To see 1.21, first observe that the right side is a constant multiple of $g\circ Y$ and hence satisfies the measurability condition that it be in σY. To check for the projection property, let V be positive and in σY; then $V = h\circ Y$ for some h in $\mathcal{D}_+$, and the required equality, namely,

$$\mathbb{E}\; f\circ X\; g\circ Y\; h\circ Y = \mathbb{E}\; h\circ Y\; g\circ Y\; \mathbb{E}\; f\circ X$$

follows from the definition of independence (see Proposition II.5.6).

1.22 *Continuation.* Let X and Y be as in 1.20 above. Then, for every positive h in $\mathcal{E} \otimes \mathcal{D}$,

1.23 $$\mathbb{E}_{\sigma Y}\ h(X,Y) = \bar{h}\circ Y,$$

where $\bar{h}(y) = \mathbb{E}\ h(X,y)$ for each y in D. First, we note that $\bar{h} \in \mathcal{D}_+$ by Fubini's theorem, and thus $\bar{h}\circ Y$ has the measurability required for it to be a conditional expectation given Y. As to the projection property, we observe that the collection of h in $\mathcal{E} \otimes \mathcal{D}$ for which

1.24 $$\mathbb{E}\ V \cdot h(X,Y) = \mathbb{E}\ V \cdot \bar{h}\circ Y$$

for some fixed V in $(\sigma Y)_+$ is a monotone class, and it includes all such h of the form $h(x,y) = f(x)g(y)$ by the preceding example. Thus, by the monotone class theorem, 1.24 holds for all positive h in $\mathcal{E} \otimes \mathcal{D}$ for which the conditional expectation is defined (that is, if $\mathbb{E}\ h(X,Y)$ exists).

1.25 *Conditional expectation of a part given the whole.* Let X and Y be independent and gamma distributed with the respective shape indices a and b, and the same scale parameter c. Then we have seen in Example II.2.11 that $Z = X + Y$ and $U = X/(X+Y)$ are independent, and some easy computations give $\mathbb{E}X = a/c$ and $\mathbb{E}Z = (a+b)/c$. Now, $\mathbb{E}\ UZ = \mathbb{E}U \cdot \mathbb{E}Z$ by independence, from which we solve for $\mathbb{E}U$ to get $\mathbb{E}\ U = a/(a+b)$ since $UZ = X$. By the same token, using 1.21, we have

$$\mathbb{E}_{\sigma Z}X = \mathbb{E}_{\sigma Z}UZ = Z\ \mathbb{E}U = \frac{a}{a+b}Z.$$

Exercises

1.26 *Relationship between $\mathbb{E}_{\mathcal{F}}X$ and $\mathbb{E}_H X$.* Let H be an event and let $\mathcal{F} = \sigma H = \{\emptyset, H, H^c, \Omega\}$. Show that $\mathbb{E}_{\mathcal{F}}X(\omega) = \mathbb{E}_H X$ for all ω in H.

1.27 *Remaining lifetime.* In Example 1.19, suppose that μ is the exponential distribution with parameter c. Show that, for $H = \{X > t\}$,

$$\mathbb{E}_H X = t + \mathbb{E}X.$$

Heuristically, then, if we know that $X(\omega) > t$ but nothing further about ω, our estimate of the lifetime $X(\omega)$ is $t + 1/c$. That is, the remaining lifetime at time t is as if the device is new at time t. This property characterizes the exponential distribution: if μ is absolutely continuous and $\mathbb{E}(X|X > t) = t + a$ for some constant a and all $t \geq 0$, then μ is the exponential distribution with parameter $c = 1/a$. Prove this.

1.28 *Conditional probabilities–elementary setup.* Let H and G be events. The conditional probability of G given H, denoted by $\mathbb{P}_H(G)$ or by $\mathbb{P}(G|H)$, is defined to be $\mathbb{E}_H 1_G$. Show that it satisfies

$$\mathbb{P}(H \cap G) = \mathbb{P}(H)\mathbb{P}_H(G).$$

If $\mathbb{P}(H) > 0$, this defines $\mathbb{P}_H(G)$ uniquely. If $\mathbb{P}(H) = 0$, then so is $\mathbb{P}(H \cap G)$, and $\mathbb{P}_H(G)$ can be taken to be any convenient number in $[0, 1]$.

1.29 *Expectations given an event.* Let $X \in \mathcal{H}_+$ and let H be an event with $\mathbb{P}(H) > 0$. Show that, similar to the result of Exercise II.2.14,

$$\mathbb{E}_H X = \int_{\mathbb{R}_+} dt \ \mathbb{P}_H\{X > t\}.$$

1.30 *Expectations given discrete variables.* Let $X \in \mathcal{H}_+$. Let Y be a random variable taking values in some countable set D. Show that, then, $\mathbb{E}_{\sigma Y} X = f \circ Y$, where f is defined by

$$f(a) = \mathbb{E}(X|Y = a) = \int_{\mathbb{R}_+} dt \ \mathbb{P}\{X > t|Y = a\}, \quad a \in D.$$

1.31 *Bounds.* If $X \le b$ for some constant b then $\mathbb{E}_{\mathcal{F}} X \le b$. If X takes values in $[a, b]$, then so does $\mathbb{E}_{\mathcal{F}} X$.

1.32 *Conditional expectation operator.* The mapping $\mathbb{E}_{\mathcal{F}} : X \mapsto \mathbb{E}_{\mathcal{F}} X$ maps $L^p(\mathcal{H})$ into $L^p(\mathcal{F})$ for every p in $[1, \infty]$. For $p = 1$ see Remark 1.6e, for $p = 2$ see Theorem 1.11, and for $p = +\infty$ see 1.31 above.

1.33 *Continuation.* If $X_n \to X$ in L^p for some p in $[1, \infty]$, then $\mathbb{E}_{\mathcal{F}} X_n \to \mathbb{E}_{\mathcal{F}} X$ in L^p with the same p. Show.

1.34 *Continuation.* If (X_n) is uniformly integrable and converges to X in probability, then $\mathbb{E}_{\mathcal{F}} X_n \to \mathbb{E}_{\mathcal{F}} X$ in L^1. Hint: see Theorem III.4.6 and the preceding exercise.

2 Conditional Probabilities and Distributions

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$. For each event H in $\mathcal{H}$,

2.1 $$\mathbb{P}_{\mathcal{F}} H = \mathbb{E}_{\mathcal{F}} 1_H$$

is called the *conditional probability of H given $\mathcal{F}$*. In more elementary settings, for events G and H, the *conditional probability of H given G* is defined to be any number $\mathbb{P}_G(H)$ in $[0, 1]$ satisfying

2.2 $$\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}_G(H);$$

of course, it is unique when $\mathbb{P}(G) > 0$.

Regular versions

Let $Q(H)$ be a version of $\mathbb{P}_{\mathcal{F}}H$ for each H in $\mathcal{H}$. We may, and do, assume that $Q(\emptyset) = 0$ and $Q(\Omega) = 1$. Of course, $Q(H)$ is a random variable in the σ-algebra $\mathcal{F}$; let $Q_\omega(H)$ denote its value at the point ω of Ω.

At first, the mapping $Q : (\omega, H) \mapsto Q_\omega(H)$ looks like a transition probability kernel from $(\Omega, \mathcal{F})$ into $(\Omega, \mathcal{H})$: By the definition of conditional expectations, the mapping $\omega \mapsto Q_\omega(H)$ is $\mathcal{F}$-measurable for each H in $\mathcal{H}$; and by the monotone convergence property of 1.8, for every disjointed sequence (H_n) in $\mathcal{H}$,

2.3
$$Q_\omega(\cup_n H_n) = \sum_n Q_\omega(H_n)$$

for almost every ω in Ω. It is this "almost" that keeps Q from being a transition kernel, and it is a serious limitation in this case.

Generally, the almost sure event Ω_h of all ω for which 2.3 holds depends on the sequence $h = (H_n)$. The set Ω_0 of all ω for which $H \mapsto Q_\omega(H)$ is a probability measure is equal to $\cap \Omega_h$, where the intersection is over all disjointed sequences h. We need Ω_0 to be almost sure before we can fix Q to become a kernel. But, usually, there are uncountably many disjointed sequences h and, hence, the intersection Ω_0 is generally a miserable object: Ω_0 might fail to belong to $\mathcal{F}$ or even to $\mathcal{H}$, and even if Ω_0 is in $\mathcal{F}$, we might have $\mathbb{P}(\Omega_0) < 1$ or even $\mathbb{P}(\Omega_0) = 0$.

Nevertheless, it is often possible to pick versions of $Q(H)$ such that $\Omega_0 = \Omega$. Such versions are highly prized.

2.4 DEFINITION. *Let $Q(H)$ be a version of $\mathbb{P}_{\mathcal{F}}H$ for every H in $\mathcal{H}$. Then $Q : (\omega, H) \mapsto Q_\omega(H)$ is said to be a* regular version *of the conditional probability $\mathbb{P}_{\mathcal{F}}$ provided that Q be a transition probability kernel from $(\Omega, \mathcal{F})$ into $(\Omega, \mathcal{H})$.*

The reason for the popularity of regular versions is the following:

2.5 PROPOSITION. *Suppose that $\mathbb{P}_{\mathcal{F}}$ has a regular version Q. Then,*

$$QX : \omega \mapsto Q_\omega X = \int_\Omega Q_\omega(d\omega') X(\omega')$$

is a version of $\mathbb{E}_{\mathcal{F}}X$ for every random variable X whose expectation exists.

Proof. It is sufficient to prove this for X in $\mathcal{H}_+$. For such X, by (Fubini's) Theorem I.6.3 applied to the transition kernel Q and function X, we see that $QX \in \mathcal{F}_+$. It is thus enough to check the projection property, namely that, for V in $\mathcal{F}_+$,

$$\mathbb{E}\, VX = \mathbb{E}\, V\, QX.$$

Fix V. For $X = 1_H$, this is immediate from the definition of $Q(H) = Q1_H$ as a version of $\mathbb{P}_{\mathcal{F}}H = \mathbb{E}_{\mathcal{F}}1_H$. This extends first to simple random variables X and then to arbitrary positive X by the linearity of, and the monotone convergence theorem for, the operators $X \mapsto QX$ and $Z \mapsto \mathbb{E}Z$. $\square$

Existence of a regular version for $\mathbb{P}_{\mathcal{F}}$ requires conditions either on $\mathcal{F}$ or on $\mathcal{H}$. For instance, if $\mathcal{F}$ is generated by a partition (G_n) of Ω, which is the case if $\mathcal{F}$ is generated by a random variable taking values in a countable space, then

2.6 $$Q_\omega(H) = \sum_n (\mathbb{P}_{G_n} H) 1_{G_n}(\omega) \ , \qquad \omega \in \Omega, H \in \mathcal{H},$$

defines a regular version, since 2.2 yields a probability measure $H \mapsto \mathbb{P}_G H$ for each G. When $\mathcal{F}$ is arbitrary, the approach is to define $Q(H)$ for all H from a judiciously chosen collection of versions $Q(H_n)$ for some sequence (H_n) in $\mathcal{H}$, and this requires conditions or limitations on $\mathcal{H}$. The following is the general result on existence. We shall give the proof afterward in Remark 2.11; see Chapter I, Section 2 for standard spaces.

2.7 THEOREM. *If $(\Omega, \mathcal{H})$ is a standard measurable space, then $\mathbb{P}_{\mathcal{F}}$ has a regular version.*

Conditional distributions

Let Y be a random variable taking values in some measurable space $(E, \mathcal{E})$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$. Then, the *conditional distribution of Y given $\mathcal{F}$* is any transition probability kernel $L : (\omega, B) \mapsto L_\omega(B)$ from $(\Omega, \mathcal{F})$ into $(E, \mathcal{E})$ such that

2.8 $$\mathbb{P}_{\mathcal{F}}\{Y \in B\} = L(B), \qquad B \in \mathcal{E}.$$

If $\mathbb{P}_{\mathcal{F}}$ has a regular version Q, then

2.9 $$L_\omega(B) = Q_\omega\{Y \in B\}, \qquad \omega \in \Omega, \ B \in \mathcal{E},$$

defines a version L of the conditional distribution of Y given $\mathcal{F}$. In general, the problem is equivalent to finding a regular version of $\mathbb{P}_{\mathcal{F}}$ restricted to the σ-algebra generated by Y. The following is the standard result.

2.10 THEOREM. *If $(E, \mathcal{E})$ is a standard measurable space, then there exists a version of the conditional distribution of Y given $\mathcal{F}$.*

2.11 REMARK. Theorem 2.7 is a straightforward corollary of the preceding theorem: Suppose that $(\Omega, \mathcal{H})$ is standard. Define $Y(\omega) = \omega$ for all ω in Ω. Then, Theorem 2.10 applies with $(E, \mathcal{E}) = (\Omega, \mathcal{H})$, and the conditional distribution of Y given $\mathcal{F}$ is precisely the regular version of $\mathbb{P}_{\mathcal{F}}$ in view of 2.8 and the fact that $H = \{Y \in H\}$ for every H in $\mathcal{H} = \mathcal{E}$.

Proof of 2.10. First, we give the proof for $E = \bar{\mathbb{R}}$ and $\mathcal{E} = \mathcal{B}(\bar{\mathbb{R}})$. For each rational number q, let

$$C_q = \mathbb{P}_{\mathcal{F}}\{Y \le q\}.$$

We shall construct the conditional distribution L of Y given $\mathcal{F}$ from these countably many random variables C_q.

Since $\{Y \le q\} \subset \{Y \le r\}$ for $q < r$, the monotonicity of conditional expectations implies that the event $\Omega_{qr} = \{C_q \le C_r\}$ in $\mathcal{F}$ is almost sure for every pair of rationals q and r with $q < r$. Let Ω_0 be the intersection of all those Ω_{qr}; it belongs to $\mathcal{F}$ and is almost sure. Fix ω in Ω_0. The mapping $q \mapsto C_q(\omega)$ from the rationals into $[0,1]$ is increasing, and thus, for each t in $\mathbb{R}$, the limit $\bar{C}_t(\omega)$ of $C_q(\omega)$ over all rationals $q > t$ exists. The resulting function $t \mapsto \bar{C}_t(\omega)$ is a cumulative distribution function on $\mathbb{R}$, and there is a unique probability measure $\bar{L}_\omega$ on $(E, \mathcal{E})$ that admits $t \mapsto \bar{C}_t(\omega)$ as its distribution function. We define

$$L_\omega(B) = 1_{\Omega_0}(\omega)\bar{L}_\omega(B) + 1_{\Omega\setminus\Omega_0}(\omega)\delta_0(B), \quad \omega \in \Omega,\ B \in \mathcal{E},$$

where δ_0 is Dirac at 0. We proceed to show that L is as desired.

a) For each ω in Ω, L_ω is a probability measure on $(E, \mathcal{E})$.

b) Let $\mathcal{D}$ be the collection of all B in $\mathcal{E}$ for which $L(B) : \omega \mapsto L_\omega(B)$ is in $\mathcal{F}_+$. It is checked easily that $\mathcal{D}$ is a d-system. Thus, in order to show that $\mathcal{D} = \mathcal{E}$ via the monotone class theorem, it is enough to show that $[-\infty, t] \in \mathcal{D}$ for every t in $\mathbb{R}$. Fix t such, let $B = [-\infty, t]$, and note that

$$L(B) = 1_{\Omega_0} \cdot \bar{C}_t + 1_{\Omega\setminus\Omega_0}\delta_0(B) = \lim_n 1_{\Omega_0} \cdot C_{r_n} + 1_{\Omega\setminus\Omega_0}\delta_0(B),$$

where (r_n) is a sequence of rationals strictly decreasing to t. On the right side, $\Omega_0 \in \mathcal{F}$ and $\Omega \setminus \Omega_0 \in \mathcal{F}$ and $\delta_0(B)$ is a constant and every C_{r_n} is in $\mathcal{F}_+$ by choice. Hence, $L(B) \in \mathcal{F}_+$ and, therefore, $B \in \mathcal{D}$. In other words, $\omega \mapsto L_\omega(B)$ is in $\mathcal{F}_+$ for every B in $\mathcal{E}$.

c) We have shown that L is a transition probability kernel from $(\Omega, \mathcal{F})$ into $(E, \mathcal{E})$. To show that it is the conditional distribution of Y given $\mathcal{F}$, there remains to show the projection property for 2.8, that is, we need to show that

2.12 $$\mathbb{P}(H \cap \{Y \in B\}) = \mathbb{E}\ 1_H L(B)$$

for H in $\mathcal{F}$ and B in $\mathcal{E}$. By the same monotone class argument as in part (b) above, it is enough to check this for $B = [-\infty, t]$ with t in $\mathbb{R}$. Fix t and let (r_n) be a sequence of rationals strictly decreasing to t. Then, by the way C_q are chosen,

2.13 $$\mathbb{P}(H \cap \{Y \in B\}) = \lim_n \mathbb{P}(H \cap \{Y \le r_n\}) = \lim_n \mathbb{E}\ 1_H C_{r_n}.$$

On the other hand, $1_H = 1_{H\cap\Omega_0}$ almost surely, and $C_{r_n}(\omega) \to \bar{C}_t(\omega) = L_\omega(B)$ for ω in Ω_0. Thus,

2.14 $$\lim \mathbb{E}\ 1_H C_{r_n} = \mathbb{E}\ 1_{H\cap\Omega_0} L(B) = \mathbb{E}\ 1_H L(B).$$

Now, putting 2.13 and 2.14 together yields 2.12 and completes the proof for the case $E = \bar{\mathbb{R}}$.

Finally, we extend the proof to the general case where $(E, \mathcal{E})$ is a standard measurable space. Then, there is an isomorphism g from E onto some Borel subset $\hat{E}$ of $[0,1]$; let $h : \hat{E} \mapsto E$ be the functional inverse of g. The preceding part applies to the real-valued random variable $g \circ Y$ and shows the existence of the conditional distribution $\hat{L} : (\omega, B) \mapsto \hat{L}_\omega(B)$ from $(\Omega, \mathcal{F})$ into $(\hat{E}, \hat{\mathcal{E}})$ for $g \circ Y$ given $\mathcal{F}$. We now put

$$L_\omega(A) = \hat{L}_\omega(h^{-1}A), \qquad \omega \in \Omega, A \in \mathcal{E}.$$

It is obvious that L is a transition probability kernel from $(\Omega, \mathcal{F})$ into $(E, \mathcal{E})$, and observing that, for H in $\mathcal{F}$ and A in $\mathcal{E}$,

$$\mathbb{P}(H \cap \{Y \in A\}) = \mathbb{P}(H \cap \{g \circ Y \in h^{-1}A\}) = \mathbb{E}\ 1_H \cdot \hat{L}(h^{-1}A) = \mathbb{E}\ 1_H L(A)$$

completes the proof that L is the conditional distribution of Y given $\mathcal{F}$. □

Disintegrations

The usual method of constructing measures over a product space was discussed in Chapter I, Section 6. Here, we treat the converse problem of disintegrating a given measure to its components. In the next subsection, we shall provide probabilistic meanings to such constructions and disintegrations. We start with a brief recall.

Let $(D, \mathcal{D})$ and $(E, \mathcal{E})$ be measurable spaces. Let μ be a probability measure on $(D, \mathcal{D})$ and let K be a transition probability kernel from $(D, \mathcal{D})$ into $(E, \mathcal{E})$. Then, according to Theorem I.6.11, the following formula for positive f in $\mathcal{D} \otimes \mathcal{E}$ defines a probability measure π on the product space $(D \times E, \mathcal{D} \otimes \mathcal{E})$:

2.15 $$\pi f = \int_{D \times E} \pi(dx, dy) f(x, y) = \int_D \mu(dx) \int_E K(x, dy) f(x, y).$$

Indeed, π is the unique measure (on the product space) that satisfies

2.16 $$\pi(A \times B) = \int_A \mu(dx) K(x, B), \qquad A \in \mathcal{D},\ B \in \mathcal{E}$$

In keeping with the short notation system mentioned in I.6.22, we write

2.17 $$\pi(dx, dy) = \mu(dx) K(x, dy), \qquad x \in D,\ y \in E,$$

to represent π defined by 2.15 and/or 2.16. This is the usual method of constructing a measure π on the product space from the measure μ and kernel K. In the next subsection, we shall give a probabilistic meaning to 2.17: if π is the joint distribution of X and Y, then $\mu(dx)$ is "the probability that X is in the small set dx" and $K(x, dy)$ is "the conditional probability that Y is in the small set dy given that X is equal to x."

The problem of disintegration is the converse to the construction described: given a probability measure π on the product space, find μ and K such that 2.17 holds (or, equivalently, 2.15 or 2.16 holds). The following is the general result; note that this is an exact converse to Theorem I.6.11 but for the condition that $(E, \mathcal{E})$ be a standard measurable space.

2.18 THEOREM. *Let π be a probability measure on the product space $(D \times E, \mathcal{D} \otimes \mathcal{E})$. Suppose that $(E, \mathcal{E})$ is standard. Then, there exist a probability measure μ on $(D, \mathcal{D})$ and a transition probability kernel K from $(D, \mathcal{D})$ into $(E, \mathcal{E})$ such that 2.17 holds.*

Proof. We cast the problem into a special case of Theorem 2.10. Let $W = D \times E$, $\mathcal{W} = \mathcal{D} \otimes \mathcal{E}$, $P = \pi$. On the probability space $(W, \mathcal{W}, P)$, define random variables X and Y by putting $X(w) = x$ and $Y(w) = y$ for $w = (x, y)$ in W. Let μ be the distribution of X, that is, $\mu(A) = \pi(A \times E), A \in \mathcal{D}$. Since Y takes values in the standard measurable space $(E, \mathcal{E})$, by Theorem 2.10, there is a regular version L of the conditional distribution of Y given $\mathcal{F} = \sigma X$. Note that $\mathcal{F}$ consists of measurable rectangles of the form $A \times E$; thus, a random variable V is in $\mathcal{F}_+$ if and only if $V(x, y) = v(x)$, free of y, for some function v in $\mathcal{D}_+$. It follows that $L_w(B) = K(X(w), B)$, where K is a transition probability kernel from $(D, \mathcal{D})$ into $(E, \mathcal{E})$. Now, the projection property for $L(B)$ yields, if $A \in \mathcal{D}$ and $B \in \mathcal{E}$, writing E for the integration under $P = \pi$,

$$\pi(A \times B) = E\ 1_A \circ X\ 1_B \circ Y = E\ 1_A \circ X\ K(X, B) = \int_D \mu(dx) 1_A(x) K(x, B).$$

This shows 2.15 for $f = 1_{A \times B}$, and the general case follows from a monotone class argument. □

Conditional distribution of Y given X

We return to the general setup of an arbitrary probability space $(\Omega, \mathcal{H}, \mathbb{P})$. Let X and Y be random variables taking values in the measurable spaces $(D, \mathcal{D})$ and $(E, \mathcal{E})$ respectively. By the *conditional distribution of Y given X* is meant the conditional distribution of Y given $\mathcal{F}$, where $\mathcal{F} = \sigma X$, the σ-algebra generated by X. The following is the description of such. Note that its condition is fulfilled at least when $(E, \mathcal{E})$ is standard; see the preceding theorem.

2.19 THEOREM. *Suppose that the joint distribution π of X and Y has the representation 2.17. Then, the kernel L defined by*

$$L_\omega(B) = K(X(\omega), B), \qquad \omega \in \Omega, B \in \mathcal{E},$$

is a version of the conditional distribution of Y given $\mathcal{F} = \sigma X$, and for every positive f in $\mathcal{D} \otimes \mathcal{E}$,

$$\mathbb{E}_{\mathcal{F}} f(X, Y) = \int_E K(X, dy) f(X, y).$$

Proof. The statement about L is immediate from Theorem 2.10 and the observation that $\mathcal{F} = \sigma X$ consists of $\mathcal{D}$-measurable functions of X. Then, by the meaning of L, if $h \in \mathcal{E}_+$,

$$\mathbb{E}_{\mathcal{F}} h(Y) = \int_E K(X, dy)h(y).$$

Thus, if $f = g \times h$ for some g in $\mathcal{D}_+$ and h in $\mathcal{E}_+$, the conditional determinism property allows $g(X)$ to come out of $\mathbb{E}_{\mathcal{F}}$, and we have

$$\mathbb{E}_{\mathcal{F}} f(X,Y) = g(X)\mathbb{E}_{\mathcal{F}} h(Y) = \int_E K(X, dy)f(X, y)$$

as claimed. Since measurable rectangles generate $\mathcal{D} \otimes \mathcal{E}$, the monotone class theorem completes the proof. □

The claims of the preceding theorem form the exact meaning of the phrase "given that $X = x$, the conditional probability that $Y \in B$ is equal to $K(x, B)$." The intuitive meaning, of course, is that transition probability kernels represent conditional probabilities, and that conditional probabilities are often the primary data in the construction of probability measures on product spaces.

Returning to the representation 2.17, which holds usually by construction or by a result like Theorem 2.18, we add that it holds trivially if X and Y are independent, and then $K(x, B) = \nu(B)$ is free of x. The following provides another construction leading to it.

Conditional densities

This is to mention a situation that is encountered often, especially in elementary probability. In the setup of the preceding subsection, suppose that the joint distribution π of X and Y has the form

2.20 $$\pi(dx, dy) = \mu_0(dx)\nu_0(dy)p(x, y), \qquad x \in D, y \in E,$$

where μ_0 and ν_0 are σ-finite measures on $(D, \mathcal{D})$ and $(E, \mathcal{E})$ respectively, and p is a positive function in $\mathcal{D} \otimes \mathcal{E}$; often, $D = E = \mathbb{R}^d$ and $\mu_0 = \nu_0 =$ Lebesgue. This π can be put in the form 2.17:

$$\pi(dx, dy) = [\mu_0(dx)m(x)][\nu_0(dy)k(x, y)],$$

where

$$m(x) = \int_E \nu_0(dy)p(x, y), \qquad k(x, y) = \begin{cases} p(x, y)/m(x) & \text{if } m(x) > 0, \\ \int_D \mu_0(dx')p(x', y) & \text{if } m(x) = 0. \end{cases}$$

Then the function $y \mapsto k(x, y)$ is called the conditional density (with respect to ν_0) of Y given that $X = x$.

2.21 EXAMPLE. Let Y and Z be independent and have the standard Gaussian distribution on $\mathbb{R}$. As an illustration of the computations discussed above, we now derive the conditional distribution of Y given $X = Y + Z$.

First, we find the joint distribution π of X and Y. To that end, we repeat that Y is standard Gaussian, whereas the conditional distribution of X given Y is Gaussian with mean Y and variance 1, because Z is independent of Y and X is the sum of Z and the "known" quantity Y given. Thus,

$$\pi(dx, dy) = dy \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \; dx \frac{1}{\sqrt{2\pi}} e^{-(x-y)^2/2}.$$

We also know that the distribution μ of $X = Y + Z$ is Gaussian with mean 0 and variance 2, that is,

$$\mu(dx) = dx \frac{1}{\sqrt{4\pi}} e^{-x^2/4}.$$

It follows that the conditional distribution $K(x, \cdot)$ of Y given $X = x$ is

$$K(x, dy) = \frac{\pi(dx, dy)}{\mu(dx)} = dy \cdot \frac{\sqrt{4\pi}}{\sqrt{2\pi}\sqrt{2\pi}} \exp\left[-\frac{y^2}{2} - \frac{(x-y)^2}{2} + \frac{x^2}{4}\right]$$

$$= dy \frac{1}{\sqrt{\pi}} \exp\left[-(y - \frac{1}{2}x)^2\right],$$

which we recognize as the Gaussian distribution with mean $x/2$ and variance $1/2$. To re-iterate, given the sum $X = Y+Z$, the conditional distribution of Y is $B \mapsto K(X, B)$, the Gaussian distribution with mean $\frac{1}{2}X$ and variance $1/2$.

Exercises

2.22 *Conditional distributions.* Let Y take values is $(E, \mathcal{E})$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$, let L be a transition probability kernel from $(\Omega, \mathcal{F})$ into $(E, \mathcal{E})$. As usual, for g in $\mathcal{E}_+$,

$$Lg(\omega) = \int_E L_\omega(dy) g(y), \qquad \omega \in \Omega.$$

Show that, if L is the conditional distribution of Y given $\mathcal{F}$, then, for every g in $\mathcal{E}_+$,

$$\mathbb{E}_\mathcal{F} \; g \circ Y = Lg.$$

2.23 *Continuation.* Let $\mathcal{F} = \sigma X$ for some random variable X with values in $(D, \mathcal{D})$. Suppose that $L_\omega(B) = K(X(\omega), B)$ for some transition probability kernel K from $(D, \mathcal{D})$ into $(E, \mathcal{E})$. Then, in the measure-kernel-function notation, show that

$$\mathbb{E} \; f \circ X \; g \circ Y = \mathbb{E} \; f \circ X \; Kg(X) = \int_D \mu(dx) f(x) \int_E K(x, dy) g(y).$$

for every f in $\mathcal{D}_+$ and g in $\mathcal{E}_+$. In particular, then,

$$\mathbb{E}_\mathcal{F} \; g \circ Y = Kg(X).$$

2.24 *Gamma variables.* Let Y and Z be independent gamma distributed variables with shape indices a and b respectively, and the same scale parameter c. Let $X = Y + Z$. Find the kernel K such that the conditional distribution of Y given X is $K(X, \cdot)$. In particular, for $a = b = 1$, show that $K(x, dy) = \frac{1}{x} dy$ for y in $(0, x)$.

2.25 *Gaussian with gamma variance.* Let X, Y, Z be as in Exercise II.2.30. Show that the conditional distribution of Z given X is $K(X, \cdot)$ where

$$K(x, dz) = dz \frac{1}{\sqrt{2\pi x}} e^{-z^2/2x}, \qquad x > 0, z \in \mathbb{R}.$$

Show that

$$\mathbb{E}_X \; e^{irZ} = e^{-r^2 X/2},$$

and that

$$\mathbb{E} \; e^{irZ} = \mathbb{E} \; \mathbb{E}_X \; e^{irZ} = \mathbb{E} \; e^{-r^2 X/2} = \left(\frac{2c}{2c + r^2} \right)^a.$$

This should explain the workings of II.2.30.

2.26 *Independence.* Let X and Y be independent and taking values in $(D, \mathcal{D})$ and $(E, \mathcal{E})$ respectively. Let $Z = h(X, Y)$ for some h in $\mathcal{D} \otimes \mathcal{E}$. Then, show that, the conditional distribution of Z given X is given as $K(X, \cdot)$ where

$$K(x, B) = \mathbb{P}\{Z \in B \mid X = x\} = \mathbb{P}\{h(x, Y) \in B\}.$$

Moral: Given that $X = x$, in most situations, we are allowed to replace X with x in our computations. Of course, to repeat the point of Example 1.22, then, $\mathbb{E}_{\sigma X} Z = \bar{h} \circ X$ where

$$\bar{h}(x) = \mathbb{E}[Z \mid X = x] = \mathbb{E} \; h(x, Y), \qquad x \in D.$$

2.27 *Stochastic process at a random time.* This is a far-fetched corollary of the preceding. Let $Y = (Y_t)_{t \in \mathbb{R}_+}$ be a stochastic process with state space $(E, \mathcal{E})$. Suppose that the mapping $(t, \omega) \mapsto Y_t(\omega)$ from $\mathbb{R}_+ \times \Omega$ into E is measurable relative to $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{H}$ and $\mathcal{E}$; this condition is fulfilled automatically if $t \mapsto Y_t(\omega)$ is right continuous for every ω, assuming that E is topological. We think of $\mathbb{R}_+$ as the time-set and of Y_t as the state of some system at the fixed time t. Now, let T be a random time, that is, a random variable taking values in $\mathbb{R}_+$. We are interested in the state of the system at that random time, namely, Y_T.

a) Show that $Y_T : \omega \mapsto Y_{T(\omega)}(\omega)$ is a random variable taking values in $(E, \mathcal{E})$.

b) Assume that T and Y are independent. Show that, for f in $\mathcal{E}_+$,

$$\mathbb{E}_{\sigma T} \; f \circ Y_T = g \circ T, \qquad \mathbb{E} \; f \circ Y_T = \mathbb{E} \; g \circ T,$$

where

$$g(t) = \mathbb{E} \; f \circ Y_t.$$

3 Conditional Independence

This is an important generalization of the concept of independence, and it is reduced to independence when the conditioning σ-algebra is trivial.

Let $\mathcal{F}, \mathcal{F}_1, \ldots, \mathcal{F}_n$ be sub-σ-algebras of $\mathcal{H}$. Then $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are said to be *conditionally independent given* $\mathcal{F}$ if

3.1 $$\mathbb{E}_\mathcal{F}\ V_1 \cdots V_n = \mathbb{E}_\mathcal{F}\ V_1 \cdots \mathbb{E}_\mathcal{F}\ V_n$$

for all positive random variables $V_1, \ldots, V_n$ in $\mathcal{F}_1, \ldots, \mathcal{F}_n$ respectively.

This definition compares to the definition II.5.1 of independence except for the substitution of $\mathbb{E}_\mathcal{F}$ for $\mathbb{E}$. Hence, all the results for independence have their counterparts for conditional independence given $\mathcal{F}$, and these counterparts are obtained by replacing $\mathbb{E}$ with $\mathbb{E}_\mathcal{F}$ throughout. Of course, if $\mathcal{F}$ is trivial, that is, if $\mathcal{F} = \{\emptyset, \Omega\}$, then $\mathbb{E}_\mathcal{F} = \mathbb{E}$ and conditional independence given $\mathcal{F}$ is the same as independence.

Heuristically, independence of $\mathcal{F}_1$ and $\mathcal{F}_2$ meant that information $\mathcal{F}_1$ is useless as far as estimating the quantities determined by $\mathcal{F}_2$. Similarly for conditional independence given $\mathcal{F}$: given the information $\mathcal{F}$, the further information provided by $\mathcal{F}_1$ is useless in estimating quantities determined by $\mathcal{F}_2$. Here is the precise version of this remark.

3.2 Proposition. *The following are equivalent:*

a) $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *are conditionally independent given* $\mathcal{F}$*.*
b) $\mathbb{E}_{\mathcal{F}\vee\mathcal{F}_1} V_2 = \mathbb{E}_\mathcal{F} V_2$ *for every positive* V_2 *in* $\mathcal{F}_2$*.*
c) $\mathbb{E}_{\mathcal{F}\vee\mathcal{F}_1} V_2 \in \mathcal{F}$ *for every positive* V_2 *in* $\mathcal{F}_2$*.*

Proof. Throughout V, V_1, V_2 are positive and in $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2$ respectively. Consider (a). It is equivalent to having

$$\mathbb{E}_\mathcal{F}\ V_1 V_2 = (\mathbb{E}_\mathcal{F}\ V_1)(\mathbb{E}_\mathcal{F}\ V_2) = \mathbb{E}_\mathcal{F}\ (V_1\ \mathbb{E}_\mathcal{F}\ V_2),$$

where the last equality is justified by the conditional determinism property. This is in turn equivalent to having, by definition,

$$\mathbb{E}\ V V_1 V_2 = \mathbb{E}\ V V_1\ \mathbb{E}_\mathcal{F}\ V_2,$$

which is equivalent to (b), since random variables of the form VV_1 generate the σ-algebra $\mathcal{F} \vee \mathcal{F}_1$. Thus, (a) $\iff$ (b). It is obvious that (b) $\implies$ (c). Conversely, if (c) holds, then

$$\mathbb{E}_{\mathcal{F}\vee\mathcal{F}_1} V_2 = \mathbb{E}_\mathcal{F}\ \mathbb{E}_{\mathcal{F}\vee\mathcal{F}_1} V_2 = \mathbb{E}_\mathcal{F}\ V_2$$

by the conditional determinism property followed by repeated conditioning. Hence, (c) $\implies$ (b). □

3.3 REMARK. The definition of conditional independence and the preceding theorem are stated in terms of positive random variables. As usual, this is because we want to avoid the trite but annoying considerations involved with arbitrary variables.

3.4 WARNING. It is possible that $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent, but fail to be conditionally independent given $\mathcal{F}$. Here is an extreme example of this state of affairs: Let X and Y be independent and identically distributed positive random variables and let $Z = X + Y$. Then, $\mathcal{F}_1 = \sigma X$ and $\mathcal{F}_2 = \sigma Y$ are independent, but they are not independent given $\mathcal{F} = \sigma Z$. In fact, in this case, $\mathbb{E}_{\mathcal{F} \vee \mathcal{F}_1} Y = Y$ whereas $\mathbb{E}_{\mathcal{F}} Y = \frac{1}{2} Z$.

Conditional independence of random variables etc.

The definition of conditional independence is extended to various settings by following the conventions for independence. For example, for an arbitrary index set T, the sub-σ-algebras $\mathcal{F}_t,\ t \in T$, are said to be conditionally independent given $\mathcal{F}$ if $\mathcal{F}_{t_1}, \ldots, \mathcal{F}_{t_n}$ are so given $\mathcal{F}$ for all integers $n \geq 2$ and choices $t_1, \ldots, t_n$ in T. Random variables $X_t,\ t \in T$, are said to be conditionally independent given $\mathcal{F}$ if the σ-algebras they generate are so given $\mathcal{F}$. If $\mathcal{F} = \sigma X$, then "given $\mathcal{F}$"is replaced by "given X". And so on.

Exercises

It will be convenient to introduce a shorthand system for conditional independence. We propose the notation $a]c[b$ for "a and b are conditionally independent given c". The notation conveys our mental image that a and b are kept apart (to act independent of each other) by the force of c between them. The arguments a, b, c are σ-algebras ordinarily, but some or all could be random variables, events, collections, etc. If c is the trivial σ-algebra, then we omit it from notation and write $a][b$, which means that a and b are independent.

Throughout the following, letters like $\mathcal{F}$ and $\mathcal{G}$ and so on are sub-σ-algebras of $\mathcal{H}$, and X and Y and so on are random variables.

3.5 *Arithmetic of conditional independence.* Show the following:

$$\begin{aligned} \mathcal{F}_1]\mathcal{F}[\mathcal{F}_2 &\iff \mathcal{F}_2]\mathcal{F}[\mathcal{F}_1 \\ &\iff \mathcal{F}_1 \vee \mathcal{F}]\mathcal{F}[\mathcal{F}_2 \iff \mathcal{F}_1]\mathcal{F}[\mathcal{F} \vee \mathcal{F}_2 \\ &\iff \mathcal{G}_1]\mathcal{F}[\mathcal{G}_2 \qquad \text{for all } \mathcal{G}_1 \subset \mathcal{F}_1 \text{ and } \mathcal{G}_2 \subset \mathcal{F}_2. \end{aligned}$$

3.6 *Proposition 3.2.* Checks for conditional independence can always be reduced to the pairwise case of Proposition 3.2: Show that $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are conditionally independent given $\mathcal{F}$ if and only if $\mathcal{F}_1 \vee \cdots \vee \mathcal{F}_k]\mathcal{F}[\mathcal{F}_{k+1}$ for $k = 1, 2, \ldots, n-1$.

3.7 *Continuation.* As a corollary to Proposition 3.2, show the following. Supposing that $\mathcal{F} \subset \mathcal{F}_1$, we have $\mathcal{F}_1]\mathcal{F}[\mathcal{F}_2$ if and only if $\mathbb{E}_{\mathcal{F}_1} V_2 \in \mathcal{F}$ for all positive V_2 in $\mathcal{F}_2$.

3.8 *Uses of the monotone class theorems.* Show that the following are equivalent,

a) $\mathcal{F}_1]\mathcal{F}[\mathcal{F}_2$.

b) $\mathbb{P}_{\mathcal{F}\vee\mathcal{F}_1} H \in \mathcal{F}_+$ for every event H in some p-system that generates $\mathcal{F}_2$.

c) $\mathbb{P}_{\mathcal{F}}(H_1 \cap H_2) = (\mathbb{P}_{\mathcal{F}} H_1)(\mathbb{P}_{\mathcal{F}} H_2)$ for every H_1 in some p-system generating $\mathcal{F}_1$ and every H_2 in some p-system generating $\mathcal{F}_2$.

3.9 *Continuation.* Show that $\mathcal{F}_1]\mathcal{F}[\{X_t : t \in T\}$ if and only if $\mathcal{F}_1]\mathcal{F}[\{X_t : t \in T'\}$ for every finite $T' \subset T$. Here, each X_t is a random variable taking values in some measurable space $(E_t, \mathcal{E}_t)$.

3.10 *Continuation.* Let $X_1, \ldots, X_n$ be random variables taking values in $(E_1, \mathcal{E}_1)$, …, $(E_n, \mathcal{E}_n)$. In view of Proposition 3.2, $\mathcal{F}_1]\mathcal{F}[(X_1, \ldots, X_n)$ if and only if $\mathbb{E}_{\mathcal{F}_1\vee\mathcal{F}} V$ is in $\mathcal{F}$ for every positive V in $\sigma(X_1, \ldots, X_n)$, that is, for every V having the form $V = f(X_1, \ldots, X_n)$ for some positive f in $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$. Show that, in fact, it is sufficient to check the condition for f having the form

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n), \qquad x_1 \in E_1, \ldots, x_n \in E_n,$$

where each f_k is in $\mathcal{E}_k$ and positive. Furthermore, each f_k can be taken to be bounded, or the indicator of an arbitrary set in $\mathcal{E}_k$, or the indicator of a set belonging to some p-system that generates $\mathcal{E}_k$, or, assuming that E_k is topological and $\mathcal{E}_k = \mathcal{B}(E_k)$, a bounded continuous function. Finally, if $E_k = \mathbb{R}^d$ with $\mathcal{E}_k = \mathcal{B}(\mathbb{R}^d)$, one can take $f_k(x) = \exp(i r_k \cdot x)$ with r_k in $\mathbb{R}^d$ and $r \cdot x$ denoting the inner product of r and x.

3.11 *Conditional independence as independence.* Suppose that there is a regular version Q of the conditional probability $\mathbb{P}_{\mathcal{F}}$. Suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent under the probability measure Q_ω, this being true for $\mathbb{P}$-almost every ω, that is, there exists an almost sure event Ω_0 such that

$$Q_\omega V_1 V_2 = (Q_\omega V_1)(Q_\omega V_2)$$

for every ω in Ω_0 and positive V_1 in $\mathcal{F}_1$ and positive V_2 in $\mathcal{F}_2$. Then, $\mathcal{F}_1]\mathcal{F}[\mathcal{F}_2$. The converse holds as well if, further, $\mathcal{F}_1$ and $\mathcal{F}_2$ are separable.

4 Construction of Probability Spaces

Our object is the construction of probability spaces and random variables corresponding to certain random experiments. We describe two basic constructions: Ionescu-Tulcea's and Kolmogorov's. Together, they show the existence of all the probability spaces that were ever needed. In particular, they yield the existence and construction of the Lebesgue measure on $(0, 1)$, and therefore the existence of all measures that were ever discussed.

To highlight the issue involved, recall from Chapter III our oft-repeated refrain, "let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space and let $X_1, X_2, \ldots$ be independent random variables with distributions $\mu_1, \mu_2, \ldots$". Do such things exist? After all, related to each random variable there are at least as many events as there are points in $\mathbb{R}$, and there are infinitely many random variables, and $\mathcal{H}$ must include all those events and their complements and all countable unions and intersections of them. Now think of the conditions on $\mathbb{P}$: it must assign a probability to each event, and do it so that countable additivity condition is fulfilled for every disjointed sequence of events. Moreover, independence of the random variables requires that certain multiplicative rules be obeyed. What if the conditions are too onerous for $\mathbb{P}$ to bear, that is, what if there can be no such $\mathbb{P}$?

The first theorem below shows, as a special case, that such things do indeed exist. Proofs are not enlightening, but the constructions leading to the theorem clarifies many of the concepts discussed earlier. In particular, note the natural appearance of conditional probabilities as primary data, as things from which $\mathbb{P}$ is constructed rather than as things derived from $\mathbb{P}$.

Construction of chains: description of data and goal

The data for the problem consist of some measurable spaces $(E_0, \mathcal{E}_0)$, $(E_1, \mathcal{E}_1), \ldots$, some probability measure μ on $(E_0, \mathcal{E}_0)$, and some transition probability kernels $K_1, K_2, \ldots$ where, for each n in $\mathbb{N}$, the kernel K_{n+1} is from

4.1 $$(F_n^o, \mathcal{F}_n^o) = (E_0 \times \cdots \times E_n, \mathcal{E}_0 \otimes \cdots \otimes \mathcal{E}_n)$$

into $(E_{n+1}, \mathcal{E}_{n+1})$.

We regard the data as follows. A random experiment is being conducted. It consists of an infinite chain of trials. The set E_n is the space of all possible outcomes of the nth trial. Our abilities to detect and discern are such that we can tell, for each A in $\mathcal{E}_n$, whether the nth trial's outcome is in A. The law governing the initial trial is described by μ; for each A in $\mathcal{E}_0$, the probability is $\mu(A)$ that the outcome of the initial trial belongs to A. Having performed the trials up to and including the nth, if the outcomes were $x_0, \ldots, x_n$ in $E_0, \ldots, E_n$ respectively, then the law governing the next trial is such that $K_{n+1}(x_0, \ldots, x_n; A)$ is the probability that the outcome belongs to the set A in $\mathcal{E}_{n+1}$.

The goal is to construct a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ that models the experiment described.

Construction and analysis

Each possible outcome of the experiment is a sequence $\omega = (x_0, x_1, \ldots)$ with x_n in E_n for each n in $\mathbb{N}$. Accordingly, we define the sample space Ω and the collection $\mathcal{H}$ of all events by

4.2 $$(\Omega, \mathcal{H}) = \otimes_{n \in \mathbb{N}} (E_n, \mathcal{E}_n).$$

We let $X_0, X_1, \ldots$ be the coordinate variables: for each n,

4.3 $$X_n(\omega) = x_n \qquad \text{if } \omega = (x_0, x_1, \ldots).$$

Obviously, X_n takes values in $(E_n, \mathcal{E}_n)$; for each outcome ω of the experiment, $X_n(\omega)$ is the result of the nth trial. Similarly, $Y_n = (X_0, \ldots, X_n)$ denotes the result of the trials up to and including the nth; it is a random variable taking values in $(F_n^o, \mathcal{F}_n^o)$ defined by 4.1.

There remains to "construct" the probability measure $\mathbb{P}$ consistent with the data. We start with the properties it should have. The interpretation we gave to $\mu, K_1, K_2, \ldots$ suggests that the distribution of Y_n be the probability measure π_n on $(F_n^o, \mathcal{F}_n^o)$ given by (see I.6.21 *et seq.*)

4.4 $$\begin{aligned}\pi_n(dx_0, \ldots, dx_n)&\\ &= \mu(dx_0)K_1(x_0, dx_1)K_2(x_0, x_1; dx_2) \cdots K_n(x_0, \ldots, x_{n-1}; dx_n).\end{aligned}$$

Let $\mathcal{F}_n = \sigma Y_n$. Every H in $\mathcal{F}_n$ has the form

4.5 $$H = \{Y_n \in B\} = B \times E_{n+1} \times \cdots, \quad B \in \mathcal{F}_n^o,$$

and then H is said to be a *cylinder with base* B in $\mathcal{F}_n^o$. Since the measure $\mathbb{P}$ being sought must yield π_n as the distribution of Y_n, we must have

4.6 $$\mathbb{P}(H) = \pi_n(B) \qquad \text{if } H \in \mathcal{F}_n,\ H \text{ has base } B \in \mathcal{F}_n^o.$$

This specifies $\mathbb{P}(H)$ for every H in $\mathcal{H}^o = \cup_n \mathcal{F}_n$ in a consistent manner: if H is a cylinder with base B in $\mathcal{F}_n^o$ and, at the same time, with base A in $\mathcal{F}_m^o$ for some $m < n$, then $B = A \times E_{m+1} \times \cdots \times E_n$ and 4.4 implies that $\pi_n(B) = \pi_m(A) = \mathbb{P}(H)$ unambiguously. There remains to show that there is $\mathbb{P}$ satisfying 4.6.

Ionescu-Tulcea's theorem

4.7 Theorem. *There exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{H})$ such that 4.6 holds.*

4.8 Remarks. a) Let $\mathbb{P}$ be as promised by the theorem. Then, 4.4 and 4.6 imply that

$$\mathbb{P}_{\mathcal{F}_m}\{X_{m+1} \in A\} = K_{m+1}(X_0, \ldots, X_m; A), \qquad A \in \mathcal{E}_{m+1}.$$

Thus the essential data $K_1, K_2, \ldots$ provide the conditional distributions of the next variable X_{m+1} given the past history $\mathcal{F}_m$ for each m in $\mathbb{N}$.

b) Fix n in $\mathbb{N}$. Fix H in $\mathcal{F}_n$. Let B be its base in $\mathcal{F}_n^o$. It follows from 4.4 and 4.6 that

$$\mathbb{P}_{\mathcal{F}_m}(H) = Q_m(X_0, \ldots, X_m; H)$$

where the kernels Q_m from $(F_m^o, \mathcal{F}_m^o)$ into $(\Omega, \mathcal{H})$ satisfy

4.9 $Q_m(x_0, \ldots, x_m; H) = 1_B(x_0, \ldots, x_n) = 1_H(x_0, x_1, \ldots)$ if $m \geq n$,

$$Q_m(x_0, \ldots, x_m; H) = \int_{E_{m+1}} K(x_0, \ldots, x_m; dx_{m+1}) Q_{m+1}(x_0, \ldots, x_{m+1}; H),$$

4.10

if $0 \leq m < n$. Moreover,

4.11 $$\mathbb{P}(H) = \int_{E_0} \mu(dx_0) Q_0(x_0; H).$$

Of these, 4.9 and 4.11 are obvious. To see 4.10, we use repeated conditioning $\mathbb{E}_{\mathcal{F}_m} = \mathbb{E}_{\mathcal{F}_m}\mathbb{E}_{\mathcal{F}_{m+1}}$ to get

$$Q_m(X_0, \ldots, X_m; H) = \mathbb{E}_{\mathcal{F}_m} Q_{m+1}(X_0, \ldots, X_{m+1}; H)$$

and evaluate the right side using part (a) above with Theorem 2.19.

Proof. a) We start by noting that $\mathcal{H}^o = \cup \mathcal{F}_n$ is an algebra and that it generates $\mathcal{H}$. In view of 4.6, the mapping $H \mapsto \mathbb{P}(H)$ from $\mathcal{F}_n$ into $[0,1]$ is a probability measure on $(\Omega, \mathcal{F}_n)$, and this is true for every n in $\mathbb{N}$. It follows that the mapping $\mathbb{P}$ from $\mathcal{H}^o$ into $[0,1]$ is finitely additive: if G and H are in $\mathcal{H}^o$ and disjoint, then there is n such that both of them belong to $\mathcal{F}_n$, and therefore, $\mathbb{P}(G \cup H) = \mathbb{P}(G) + \mathbb{P}(H)$. We shall show below that

4.12 $$(H_k) \subset \mathcal{H}^o \text{ and } H_k \searrow \emptyset \quad \Longrightarrow \quad \mathbb{P}(H_k) \searrow 0.$$

Once this is shown, the existence of $\mathbb{P}$ on $(\Omega, \mathcal{H})$ satisfying 4.6 for all n will follow from the standard theorems on extensions of measures from algebras to σ-algebras (see Caratheodory's theorem, I.3.19). Uniqueness of $\mathbb{P}$ is immediate from Proposition I.3.7 since $\mathcal{H}^o$ is a p-system generating $\mathcal{H}$.

b) Each H in $\mathcal{H}^o$ is a cylinder with some base B in $\mathcal{F}_n^o$ for some n in $\mathbb{N}$; and, if so, we define $Q_m(x_0, \ldots, x_m; H)$ for m in $\mathbb{N}$ and $(x_0, \ldots, x_m)$ in F_m^o starting with 4.9 and continuing with 4.10 iteratively downward, and observe that 4.11 holds and that 4.10 holds for all $m < n$ and $m \geq n$.

To show 4.12, pick (H_k) as described. Then $\mathbb{P}(H_k)$ is well-defined for each k and decreases with k, since $\mathbb{P}$ is finitely additive on $\mathcal{H}^o$. Suppose for the moment that $\lim_k \mathbb{P}(H_k) > 0$; we shall show that this leads to a contradiction.

Replace H in 4.11 by H_k, take limits as $k \longrightarrow \infty$ on both sides, and pass the limit on the right side inside the integral with an appeal to the bounded convergence theorem. The limit on the left side is strictly positive by assumption; so there must exist x_0^* in E_0 such that $a_0 = \lim_k Q_0(x_0^*, H_k)$ is strictly positive. We now make the induction hypothesis that there exists $(x_0^*, \ldots, x_m^*)$ in F_m^o such that

4.13 $$a_m = \lim_k Q_m(x_0^*, \ldots, x_m^*; H_k) > 0.$$

Recall that 4.10 holds for all m, and by the bounded convergence theorem,

$$a_m = \int_{E_{m+1}} K_{m+1}(x_0^*, \ldots, x_m^*; dx_{m+1}) \lim_k Q_{m+1}(x_0^*, \ldots, x_m^*, x_{m+1}; H_k).$$

Since $a_m > 0$, there must exist x_{m+1}^* in E_{m+1} such that 4.13 holds for $m+1$ as well. Thus, in view of 4.9, we have shown the existence of a sequence $\omega^* = (x_0^*, x_1^*, \ldots)$ in Ω such that $\lim_k 1_{H_k}(\omega^*)$ is strictly positive and, therefore, is equal to 1. This means that $\omega^* \in H_k$ for all k large enough and, thus, for all k since (H_k) is decreasing. In other words, $w^* \in \cap H_k$, which contradicts the hypothesis that $H_k \searrow \emptyset$. Hence, we must have $\lim_k \mathbb{P}(H_k) = 0$ as needed to complete the proof of 4.7. □

Initial distribution

It is often desirable to treat the initial distribution μ as a variable rather than as a part of the given data. To that end, it is customary to write $\mathbb{P}^\mu$ for the probability measure $\mathbb{P}$ of the preceding theorem. Of course, to each probability measure μ on $(E_0, \mathcal{E}_0)$ there corresponds a unique probability measure $\mathbb{P}^\mu$ on $(\Omega, \mathcal{H})$. In particular, let $\mathbb{P}^x$ denote $\mathbb{P}^\mu$ when $\mu = \delta_x$, Dirac measure sitting at the point x in E_0. It is clear from 4.4 and 4.6 that $x \mapsto \mathbb{P}^x(H)$ is $\mathcal{E}_0$-measurable for each H in $\mathcal{H}$. Thus, $(x, H) \mapsto \mathbb{P}^x(H)$ is a transition probability kernel from $(E_0, \mathcal{E}_0)$ into $(\Omega, \mathcal{H})$ and

4.14 $$\mathbb{P}^\mu(H) = \int_{E_0} \mu(dx)\mathbb{P}^x(H), \qquad H \in \mathcal{H},$$

for every measure μ on $(E_0, \mathcal{E}_0)$. For fixed x in E_0, the measure $\mathbb{P}^x$ is called the *probability law* of the chain $X = (X_n)$ started at x, since

4.15 $$\mathbb{P}^x\{X_0 = x\} = 1$$

provided that the singleton $\{x\}$ belong to $\mathcal{E}_0$ to ensure that $\{X_0 = x\}$ is an event. It can be also regarded as the conditional law of X given that $X_0 = x$.

Kolmogorov extension theorem

As opposed to a chain, we now consider the question of existence for a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ that can carry a process $\{X_t : t \in I\}$ where the index set I is arbitrary. In this case, Kolmogorov's theorem is the most general known, but requires that all the X_t take values in the same space $(E, \mathcal{E})$ and that $(E, \mathcal{E})$ be a standard measurable space.

Let I be an arbitrary index set. Let $(E, \mathcal{E})$ be a measurable space. The data are the probability measures π_J, one for each finite subset J of I, on the product space $(E^J, \mathcal{E}^J)$. The goal is to construct a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and a stochastic process $X = (X_t)_{t \in I}$ over it such that π_J is the distribution of the random variable $X_J = (X_t)_{t \in J}$ for each finite $J \subset I$.

We start by letting $(\Omega, \mathcal{H}) = (E, \mathcal{E})^I$, that is, Ω is the collection of all functions $t \mapsto \omega(t)$ from I into E, and $\mathcal{H}$ is the σ-algebra generated by the finite-dimensional measurable rectangles. We define the X_t to be the coordinate variables; that is, $X_t(\omega) = \omega(t)$ for all t and ω. Obviously, each X_t is measurable with respect to $\mathcal{H}$ and $\mathcal{E}$, and in fact $\mathcal{H} = \sigma\{X_t : t \in I\}$.

For $I \supset J \supset K$, we let p_{JK} denote the natural projection from E^J onto E^K. For instance, if $J = (s, t, u)$ and $K = (u, t)$, then $p_{IJ}(\omega) = (\omega(s), \omega(t), \omega(u))$ and $p_{JK}(x, y, z) = (z, y)$ and $p_{IK}(\omega) = (\omega(u), \omega(t))$. We let $\mathcal{I}_f$ denote the collection of all finite sequences of elements of I, and $\mathcal{I}_c$ the collection of all infinite (countable) sequences.

The probability measure $\mathbb{P}$ we are seeking will be the probability law of X; accordingly, we want

4.16 $$\mathbb{P}\{X_J \in A\} = \pi_J(A), \qquad A \in \mathcal{E}^J,\ J \in \mathcal{I}_f.$$

This requires that the finite-dimensional distributions be consistent:

4.17 $$\pi_K = \pi_J \circ p_{JK}^{-1}, \qquad K \subset J \in \mathcal{I}_f,$$

since $X_K = p_{JK} \circ X_J$ for $K \subset J$. The following is the Kolmogorov extension theorem.

4.18 THEOREM. *Suppose that $(E, \mathcal{E})$ is a standard measurable space and that $\{\pi_J : J \in \mathcal{I}_f\}$ satisfies the consistency condition 4.17. Then, there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{H})$ such that 4.16 holds.*

Proof. We start by constructing a probability measure P_J on $(E, \mathcal{E})^J$ for every J in $\mathcal{I}_c$. Fix J so, say, $J = (t_0, t_1, \ldots)$, and let $J_n = (t_0, \ldots, t_n)$ and $\hat{\pi}_n = \pi_{J_n}$. Observe that the $\hat{\pi}_n$ have the representation 4.4: this is trivial for $n = 0$, and assuming it is true for n, it follows from the disintegration theorem 2.18 with $(D, \mathcal{D}) = (E, \mathcal{E})^{J_n}$ that it is true for $n + 1$ as well (this is where the standardness of $(E, \mathcal{E})$ gets used). Thus, Ionescu-Tulcea's theorem applies to show that there exists a probability measure P_J on $(E, \mathcal{E})^J$ such that the image of P_J under p_{JJ_n} is $\hat{\pi}_n$ for each n.

In fact, for $K \subset J \in \mathcal{I}_c$, the probability measures P_K and $P_J \circ p_{JK}^{-1}$ coincide over the finite-dimensional cylinders in $\mathcal{E}^K$ with bases in $\mathcal{E}^L$ for finite subsets L of K, and hence, since such cylinders generate $\mathcal{E}^K$, we have $P_K = P_J \circ p_{JK}^{-1}$.

By Proposition II.4.6, for every H in $\mathcal{H}$ there is J in $\mathcal{I}_c$ such that $H = \{X_J \in A\}$ for some A in $\mathcal{E}^J$; then we put $\mathbb{P}(H) = P_J(A)$. This assignment is without ambiguities: if $H = \{X_J \in A\} = \{X_K \in B\}$ for A in $\mathcal{E}^J$ and B in $\mathcal{E}^K$ for some J and K in $\mathcal{I}_c$, then $L = J \cup K$ is in $\mathcal{I}_c$, and

$$P_J(A) = P_L(p_{LJ}^{-1} A) = P_L(p_{LK}^{-1} B) = P_K(B).$$

We now show that $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{H})$. Let $(H_n) \subset \mathcal{H}$ be disjointed and have union H. For each n, there is J_n in $\mathcal{I}_c$ such that $H_n = \{X_{J_n} \in A_n\}$ for some A_n belonging to $\mathcal{E}^{J_n}$. We may assume that the

J_n are all the same by replacing J_n with $J = \cup J_n$; and, then, $H = \{X_J \in A\}$ with $A = \cup A_n$, the A_n being disjoint. Hence, by the countable additivity of P_J,

$$\mathbb{P}(H) = P_J(A) = \sum_n P_J(A_n) = \sum_n \mathbb{P}(H_n).$$

So, $\mathbb{P}$ is countably additive and $\mathbb{P}(\Omega) = 1$ obviously. Finally, $\mathbb{P}$ is the unique probability measure satisfying 4.16 since the cylinders $\{X_J \in A\}$ form a p-system that generates $\mathcal{H}$. □

5 Special Constructions

This section is devoted to special cases of the probability spaces constructed in the preceding section, as well as certain alternatives to such constructions. The setup and notations are carried over from the preceding section.

Independent sequences

In the setup for chains, suppose that the transition kernels K_n have the following form: for each n in $\mathbb{N}$ there is a probability measure μ_n on $(E_n, \mathcal{E}_n)$ such that $\mu = \mu_0$ and

5.1 $$K_{n+1}(x_0, \ldots, x_n; A) = \mu_{n+1}(A), \quad A \in \mathcal{E}_{n+1}.$$

Heuristically, this corresponds to the situation where the law governing the $(n+1)^{\text{th}}$ trial is independent of the results $x_0, \ldots, x_n$ of the previous trials. Then, 4.4 becomes

5.2 $$\pi_n(dx_0, \ldots, dx_n) = \mu_0(dx_0)\mu_1(dx_1) \cdots \mu_n(dx_n),$$

that is, π_n is the product of $\mu_0, \ldots, \mu_n$. Equivalently, then, the random variables $X_0, \ldots, X_n$ are independent for each n, and hence, $(X_n)_{n\in\mathbb{N}}$ is an independency. In this case, the probability space $(\Omega, \mathcal{H}, \mathbb{P})$ is said to be the *product* of the probability spaces $(E_n, \mathcal{E}_n, \mu_n)$, $n \in \mathbb{N}$, and the following notation is used to express it:

5.3 $$(\Omega, \mathcal{H}, \mathbb{P}) = \otimes_{n\in\mathbb{N}}(E_n, \mathcal{E}_n, \mu_n).$$

Variations where $\mathbb{N}$ is replaced by $\mathbb{N}^* = \{1, 2, \ldots\}$ or by $\mathbb{Z} = \{\ldots, -1, 0, 1, 2, \ldots\}$ or by some other countable set are self-explanatory.

A further special case is where the spaces $(E_n, \mathcal{E}_n, \mu_n)$ are the same for all n, that is, say

5.4 $$(E_n, \mathcal{E}_n, \mu_n) = (E, \mathcal{E}, \mu), \quad n \in \mathbb{N}.$$

Then, instead of 5.3, it is usual to write

5.5 $$(\Omega, \mathcal{H}, \mathbb{P}) = (E, \mathcal{E}, \mu)^{\mathbb{N}}.$$

In this case, the coordinate variables $X_0, X_1, \ldots$ take values in the same measurable space, and, in addition to being independent, they are identically distributed.

Markov chains

We start with the most useful and the most common case: This is where all the trials are on the same space, that is,

5.6 $$(E_n, \mathcal{E}_n) = (E, \mathcal{E}), \quad n \in \mathbb{N},$$

and there is a Markov kernel P on $(E, \mathcal{E})$, that is, a transition probability kernel P from $(E, \mathcal{E})$into $(E, \mathcal{E})$, such that

5.7 $$K_{n+1}(x_0, \ldots, x_n; A) = P(x_n, A)$$

for all n in $\mathbb{N}$ and $x_0, \ldots, x_n$ in E and A in $\mathcal{E}$. In other words, the probability law governing the $(n+1)^{\text{th}}$ trial is independent of n and of all the previous results $x_0, \ldots, x_{n-1}$ except the result x_n of the nth trial. Then, $X = (X_n)_{n \in \mathbb{N}}$ is said to be a *Markov chain* over $(\Omega, \mathcal{H}, \mathbb{P})$ with state space $(E, \mathcal{E})$ and initial distribution μ and transition kernel P.

Moreover, the kernel P is considered to be the only ingredient defining the law of X, and μ is treated as a variable by utilizing the notation $\mathbb{P}^\mu$ and $\mathbb{P}^x$ mentioned around 4.14.

The term "Markov chain" refers to the property, obvious from 5.7, that

5.8 $$\mathbb{P}\{X_{n+1} \in A \mid \mathcal{F}_n\} = P(X_n, A), \quad A \in \mathcal{E},$$

which displays $P(x, A)$ as the conditional probability that $X_{n+1} \in A$ given that $X_n = x$. Note the independence of this probability from n; if this last point needs to be emphasized, then X is said to be a *time-homogeneous* Markov chain.

5.9 *Non-canonical Markov chains.* Assuming 5.6 and 5.7, the construction given in the preceding section for $\Omega, \mathcal{H}, (\mathcal{F}_n), (X_n), \mathbb{P}$ is called the canonical setting of a Markov chain X with state space $(E, \mathcal{E})$, initial distribution μ, and transition kernel P. More generally, given a chain $X = (X_n)_{n \in \mathbb{N}}$ defined over some probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and having some state space $(E, \mathcal{E})$, and given some filtration $(\mathcal{F}_n)$ on $(\Omega, \mathcal{H})$ such that X_n is measurable with respect to $\mathcal{F}_n$ and $\mathcal{E}$ for each n, the chain X is said to be a time-homogeneous Markov chain with transition kernel P provided that 5.8 hold for each n in $\mathbb{N}$. We shall see several such non-canonical examples below.

5.10 *Time-inhomogeneous Markov chains.* Returning to the constructions of the previous section, assume 5.6 and replace 5.7 with

5.11 $$K_{n+1}(x_0, \ldots, x_n; A) = P_{n+1}(x_n, A),$$

where $P_1, P_2, \ldots$ are Markov kernels on $(E, \mathcal{E})$. Then, we obtain the canonical construction for a time-inhomogeneous Markov chain X. Such chains can be made time-homogeneous by incorporating time into the state space: Let $\hat{E} = \mathbb{N} \times E, \hat{\mathcal{E}} = 2^{\mathbb{N}} \otimes \mathcal{E}$, and define the Markov kernel $\hat{P}$ on $(\hat{E}, \hat{\mathcal{E}})$ such that

5.12 $$\hat{P}(y, B) = P_{n+1}(x, A) \quad \text{if } y = (n, x),\ B = \{n+1\} \times A.$$

Then, putting $\hat{X}_n = (n, X_n)$, we note that $\hat{X}_n$ is measurable with respect to $\mathcal{F}_n$ and $\hat{\mathcal{E}}$, and the Markov property holds just as in 5.8:

$$\mathbb{P}\{\hat{X}_{n+1} \in B \mid \mathcal{F}_n\} = \hat{P}(\hat{X}_n, B).$$

Thus, $\hat{X} = (\hat{X}_n)$ is a time-homogeneous Markov chain over $(\Omega, \mathcal{H}, \mathbb{P})$ with state space $(\hat{E}, \hat{\mathcal{E}})$ and transition kernel $\hat{P}$ in the sense of 5.9 above. Note that this is not the canonical construction for $\hat{X}$.

5.13 *Periodicity in time.* This is a special case of time-inhomogeneity which is encountered often in applied work where seasonal effects or the day-of-the-week effects and the like affect the transition probabilities. For instance, if X_n is to denote the inventory level for some item at the end of day n, then we would expect $P_n(x, A)$ to depend on n only through whether n is a Monday or Tuesday and so on. In such cases, there is an integer $d \geq 2$ such that the sequence of transition kernels P_n in 5.11 has the form

5.14 $$(P_1, P_2, \ldots) = (P_1, P_2, \ldots, P_d, P_1, P_2, \ldots, P_d, \ldots).$$

The corresponding time-inhomogeneous chain X can be rendered time-homogeneous by incorporating periodicity into the state space: Let $D = \{1, 2, \ldots, d\}, \mathcal{D} = 2^D, \hat{E} = D \times E, \hat{\mathcal{E}} = \mathcal{D} \otimes \mathcal{E}$, and define $\hat{P}$ as a Markov kernel on $(\hat{E}, \hat{\mathcal{E}})$ that satisfies

$$\hat{P}(y, B) = P_j(x, A) \quad \text{if } y = (i, x),\ B = \{j\} \times A,\ j = (1+i) \text{ modulo } d.$$

Then, $\hat{X}_n = (n \text{ modulo } d, X_n),\ n \in \mathbb{N}$, form a time-homogeneous Markov chain with state space $(\hat{E}, \hat{\mathcal{E}})$ and transition kernel $\hat{P}$. Again, this is not canonical.

5.15 *Cyclic Markov chains.* The concept is similar to periodicity but with the added twist that the space changes with time. Suppose X is a chain constructed as in the preceding section, but with

$$(E_0, E_1, \ldots) = (E_d, E_1, E_2, \ldots, E_d, E_1, E_2, \ldots, E_d, \ldots),$$

and similarly for the $\mathcal{E}_n$, and the conditions 5.11 and 5.14– note that P_j is a kernel from $(E_{j-1}, \mathcal{E}_{j-1})$ into $(E_j, \mathcal{E}_j), j = 1, \ldots, d$. Then, X is called a cyclic Markov chain in canonical form: if we think of X_n as the position of a particle after its nth step, the particle moves from a point in space E_d into a random

point in E_1, and then to a random point in E_2, ..., and so on cyclically, with transition probabilities depending only on the space to be visited next. The chain X can be made time-homogeneous by the technique used on periodic chains but with a more careful construction for $(\hat{E}, \hat{\mathcal{E}})$.

5.16 *k-dependent Markov chains.* For the Markov chains introduced so far, the conditional distribution of X_{n+1} given all the past $\mathcal{F}_n$ depended only on the last state X_n. In some applications, for instance if X_n is to denote the weather on day n, it is desired to make the dependence on the past a little deeper: For fixed integer $k \geq 2$, suppose that

5.17 $$K_{n+1}(x_0, \ldots, x_n; A) = P(x_{n-k+1}, \ldots, x_n; A)$$

for each $n \geq k$ for all $x_0, \ldots, x_n$ in E and A in $\mathcal{E}$, where P is a transition probability kernel from $(E, \mathcal{E})^k$ to $(E, \mathcal{E})$. Then, $X = (X_n)$ is said to be a k-dependent time-homogeneous Markov chain with state space $(E, \mathcal{E})$.

Theoretically, such chains are no different from ordinary Markov chains. To convert such a chain to an ordinary one-dependent chain, one needs to re-define the term "state": Put $(\hat{E}, \hat{\mathcal{E}}) = (E, \mathcal{E})^k$, let $\hat{X}_n = (X_n, X_{n+1}, \ldots, X_{n+k-1})$, and define the Markov kernel $\hat{P}$ on $(\hat{E}, \hat{\mathcal{E}})$ such that

5.18 $$\hat{P}(y, B) = P(y_1, \ldots, y_k; A) \quad \text{if } B = \{y_2\} \times \cdots \times \{y_k\} \times A.$$

Then, $(\hat{X}_n)$ is an ordinary Markov chain with state space $(\hat{E}, \hat{\mathcal{E}})$ and transition kernel $\hat{P}$.

Markov processes, continuous time

These are the continuous time versions of Markov chains, that is, the parameter set is $\mathbb{R}_+$. We discuss only the most fundamental cases. In the setting of Kolmogorov extension theorem of the preceding section, let the index set be $I = \mathbb{R}_+$, let μ be a probability measure on $(E, \mathcal{E})$, and, for every pair (s, t) of times in $\mathbb{R}_+$ with $s < t$, let $P_{s,t}$ be a Markov kernel on $(E, \mathcal{E})$. We are to interpret μ as the distribution of X_0, and $P_{s,t}(x, A)$ as the conditional probability that $X_t \in A$ given that $X_s = x$. Accordingly, if $J = (t_0, t_1, \ldots, t_n)$ with $0 = t_0 < t_1 < \cdots < t_n$, then we specify the distribution π_J of $X_J = (X_{t_0}, X_{t_1}, \ldots, X_{t_n})$ by

5.19 $$\begin{aligned}&\pi_J(dx_0, dx_1, dx_2, \ldots, dx_{n-1}, dx_n) \\ &= \mu(dx_0) P_{t_0,t_1}(x_0, dx_1) P_{t_1,t_2}(x_1, dx_2) \cdots P_{t_{n-1},t_n}(x_{n-1}, dx_n).\end{aligned}$$

Assuming that

5.20 $$P_{s,t} P_{t,u} = P_{s,u}, \quad 0 \leq s < t < u < \infty,$$

that is, in more explicit notation,

5.21 $$\int_E P_{s,t}(x, dy) P_{t,u}(y, B) = P_{s,u}(x, B), \quad x \in E, B \in \mathcal{E},$$

the consistency requirement 4.17 is satisfied. Then, according to Theorem 4.18, there exists a unique probability measure $\mathbb{P}^\mu$ on $(\Omega, \mathcal{H}) = (E, \mathcal{E})^{\mathbb{R}_+}$ such that 5.19 is the distribution of $(X_{t_0}, X_{t_1}, \ldots, X_{t_n})$, where $0 = t_0 < t_1 < \cdots < t_n$.

Let $\mathcal{F}_t$ be the σ-algebra generated by $\{X_s : 0 \le s \le t\}$, that is, the σ-algebra generated by the finite-dimensional cylinders of the form $\{X_{s_1} \in A_1, \ldots, X_{s_n} \in A_n\}$ with $0 \le s_1 < \cdots < s_n \le t$ and $A_1, \ldots, A_n$ in $\mathcal{E}$. Then, it follows from 5.19 and theorems on conditional expectations that

5.22 $$\mathbb{P}^\mu\{X_t \in A \mid \mathcal{F}_s\} = P_{s,t}(X_s, A), \quad A \in \mathcal{E},$$

for all $0 \le s < t$; and, of course,

5.23 $$\mathbb{P}^\mu\{X_0 \in A\} = \mu(A), \quad A \in \mathcal{E}.$$

For these reasons, $X = (X_t)_{t \in \mathbb{R}_+}$ is called a *Markov process* with state space $(E, \mathcal{E})$ and transition kernels $(P_{s,t})_{0 \le s < t < \infty}$. This is the time-inhomogeneous case.

Mathematically more interesting is the *time-homogeneous* case where $P_{s,t}$ depends on s and t only through $t - s$. Thus, replacing $P_{s,t}$ with P_{t-s}, we see that 5.20 becomes

5.24 $$P_s P_t = P_{s+t}, \quad s, t \in \mathbb{R}_+,$$

and the collection $(P_t)_{t \in \mathbb{R}_+}$ is called the *transition semigroup* of the (time-homogeneous) Markov process $X = (X_t)_{t \in \mathbb{R}_+}$ with state space $(E, \mathcal{E})$. Incidentally, the semigroup property 5.24 as well as the parental 5.20 are called *Chapman-Kolmogorov equations.*

Random fields

Returning to Kolmogorov extension theorem 4.18, suppose that the index set I is $\mathbb{R}^d$ for some fixed dimension $d \ge 1$, and suppose that the state space $(E, \mathcal{E})$ is replaced with $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. The index set can no longer be thought as time, and we now write x for the generic element of $I = \mathbb{R}^d$. Assuming that Kolmogorov's theorem applies, we obtain a stochastic process $X = \{X(x) : x \in \mathbb{R}^d\}$, and we may view it as a random field on $\mathbb{R}^d$, that is, to each x in $\mathbb{R}^d$ we associate a real-valued random variable $X(x)$.

In the further case where the state space $(E, \mathcal{E})$ is taken to be $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, the random variable $X(x)$ can be thought as the Eulerian velocity vector at x and the process X becomes a *random velocity field.* In such applications, one generally requires some smoothness (like continuity or differentiability) from the mapping $x \mapsto X(x, \omega)$ for each ω. Theorem 4.18 guarantees no such smoothness.

Chapter V

Martingales and Stochastics

This chapter is to introduce the vocabulary for describing the evolution of random systems over time. It will also cover the basic results of classical martingale theory and mention some basic processes such as Markov chains, Poisson processes, and Brownian motion. This chapter should be treated as a reference source for chapters to come.

We start with generalities on filtrations and stopping times, go on to martingales in discrete time, and then to finer results on martingales and filtrations in continuous time. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a fixed probability space in the background, and all stochastic processes are indexed by some set $\mathbb{T}$, which is either $\mathbb{N} = \{0, 1, \ldots\}$ or $\mathbb{R}_+ = [0, \infty)$ or some other subset of $\bar{\mathbb{R}} = [-\infty, +\infty]$. We think of $\mathbb{T}$ as the time-set; its elements are called times. On a first reading, the reader should take $\mathbb{T} = \mathbb{N}$.

1 Filtrations and Stopping Times

Let $\mathbb{T}$ be a subset of $\bar{\mathbb{R}}$. A *filtration* on $\mathbb{T}$ is an increasing family of sub-σ-algebras of $\mathcal{H}$ indexed by $\mathbb{T}$; that is, $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ is a *filtration* if each $\mathcal{F}_t$ is a σ-algebra on Ω, each $\mathcal{F}_t$ is a subset of $\mathcal{H}$, and $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s < t$. Given a stochastic process $X = (X_t)_{t \in \mathbb{T}}$, letting $\mathcal{F}_t = \sigma\{X_s : s \le t\}$ for each time t, we obtain a filtration $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$; it is called the *filtration generated* by X.

Heuristically, we think of a filtration $\mathcal{F}$ as a flow of information, with $\mathcal{F}_t$ representing the body of information accumulated by time t by some observer of the ongoing experiment modeled by $(\Omega, \mathcal{H}, \mathbb{P})$. Or, we may think of $\mathcal{F}_t$ as the collection of $\bar{\mathbb{R}}$-valued random variables V such that the observer can tell the value $V(\omega)$ at the latest by time t, whatever the outcome ω turns out to be. Of course, it is possible to have different observers with different

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_5,

information flows. Given two filtrations $\mathcal{F}$ and $\mathcal{G}$, we say that $\mathcal{F}$ is *finer* than $\mathcal{G}$, or that $\mathcal{G}$ is *coarser* than $\mathcal{F}$, if $\mathcal{F}_t \supset \mathcal{G}_t$ for every time t.

Adaptedness

Let $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ be a filtration. Let $X = (X_t)_{t\in\mathbb{T}}$ be a stochastic process with some state space $(E, \mathcal{E})$. Then X is said to be *adapted* to $\mathcal{F}$ if, for every time t, the variable X_t is measurable with respect to $\mathcal{F}_t$ and $\mathcal{E}$. Since $\mathcal{F}$ is increasing, this is equivalent to saying that, for each t, the numerical random variables $f \circ X_s$ belong to $\mathcal{F}_t$ for all f in $\mathcal{E}$ and all times $s \le t$.

Every stochastic process is automatically adapted to the filtration it generates. Thus, if $\mathcal{G}$ is the filtration generated by X, saying that X is adapted to $\mathcal{F}$ is the same as saying that $\mathcal{F}$ is finer than $\mathcal{G}$.

Stopping times

1.1 Definition. *Let $\mathcal{F}$ be a filtration on $\mathbb{T}$. A random time $T : \Omega \mapsto \bar{\mathbb{T}} = \mathbb{T} \cup \{+\infty\}$ is called a* stopping time *of $\mathcal{F}$ if*

1.2 $$\{T \le t\} \in \mathcal{F}_t \quad \textit{for each} \quad t \in \mathbb{T}.$$

1.3 Remarks. The condition 1.2 is equivalent to requiring that the process

1.4 $$Z_t = 1_{\{T\le t\}}, \quad t \in \mathbb{T},$$

be adapted to $\mathcal{F}$. When $\mathbb{T}$ is $\mathbb{N}$ or $\bar{\mathbb{N}}$, this is equivalent to requiring that

1.5 $$\hat{Z}_n = 1_{\{T=n\}}, \quad n \in \mathbb{N},$$

be adapted to $(\mathcal{F}_n)$; this follows from the preceding remark by noting that $\hat{Z}_n = Z_n - Z_{n-1}$.

Heuristically, a random time signals the occurrence of some physical event. The process Z defined by 1.4 is indeed the indicator of whether that event has or has not occurred: $Z_t(\omega) = 0$ if $t < T(\omega)$ and $Z_t(\omega) = 1$ if $t \ge T(\omega)$. Recalling the heuristic meaning of adaptedness, we conclude that T is a stopping time of $\mathcal{F}$ if the information flow $\mathcal{F}$ enables us to detect the occurrence of that physical event as soon as it occurs, as opposed to inferring its occurrence sometime later. In still other words, T is a stopping time of $\mathcal{F}$ if the information flow $\mathcal{F}$ is such that we can tell what $T(\omega)$ is at the time $T(\omega)$, rather than by inference at some time after $T(\omega)$. These heuristic remarks are more transparent when the time set is $\mathbb{N}$.

The following mental test incorporates all these remarks into a virtual alarm system. Imagine a computer that is being fed the flow $\mathcal{F}$ of information and that is capable of checking, at each time t, whether $\omega \in H$ for every possible ω in Ω and every event H in $\mathcal{F}_t$. If it is possible to attach to it an

alarm system that sounds exactly at time T, and only at time T, then T is a stopping time of $\mathcal{F}$. This alarm test will be heard on and off below.

1.6 EXAMPLE. Let $\mathbb{T} = \mathbb{N}$, let $\mathcal{F}$ be a filtration on $\mathbb{N}$, and let X be a process with index set $\mathbb{N}$ and some state space $(E, \mathcal{E})$. Suppose that X is adapted to $\mathcal{F}$. For fixed A in $\mathcal{E}$, let

$$T(\omega) = \inf\{\ n \in \mathbb{N} : X_n(\omega) \in A\ \}, \quad \omega \in \Omega.$$

Then T is called the time of first entrance to A. (Note that $T(\omega) = +\infty$ if $X_n(\omega)$ is never in A, which is the reason for allowing $+\infty$ as a value for random times in general.) This T is a stopping time of $\mathcal{F}$: Heuristically, X is adapted to $\mathcal{F}$ means that the computer is able to check, at each time n, whether $X_n(\omega) \in A$; and it seems trivial to design an alarm system that sounds exactly at the first n such that $X_n(\omega) \in A$. More precisely, T is a stopping time because, for each n in $\mathbb{N}$,

$$\{\ T \leq n\ \} = \bigcup_{k=0}^{n} \{\ X_k \in A\ \}$$

belongs to $\mathcal{F}_n$, since the events $\{X_k \in A\}$, $0 \leq k \leq n$, are all in $\mathcal{F}_n$. In contrast,

$$L(\omega) = 0 \vee \sup\{\ n \leq 5\ :\ X_n(\omega) \in A\ \}$$

is not a stopping time (except in some special cases depending on A, for instance, if entering A means never coming back to A). Because, if the outcome ω is such that $X_4(\omega) \in A$ and $X_5(\omega) \notin A$, then $L(\omega) = 4$, but the information we had at time 4 is not sufficient to conclude that $L(\omega) = 4$. So, there can be no alarm system that will sound at exactly $L(\omega)$.

1.7 EXAMPLE. *Counting Processes.* Let $0 < T_1 < T_2 < \cdots$ be some random times taking values in $\mathbb{R}_+$ and assume that $\lim T_n = +\infty$. Define

$$N_t = \sum_{1}^{\infty} 1_{[0,t]} \circ T_n, \quad t \in \mathbb{R}_+,$$

and note that $t \mapsto N_t$ is increasing and right-continuous and increases only by jumps of size one, and $N_0 = 0$ and $N_t < \infty$ for every t in $\mathbb{R}_+$, with $\lim_{t\to\infty} N_t = +\infty$. We may regard $T_1, T_2, \ldots$ as the times of successive arrivals at a store; then N_t is the number of arrivals during $[0, t]$. Let $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be the filtration generated by $N = (N_t)$. Then, for each integer $k \geq 1$, the time T_k is a stopping time of $\mathcal{F}$: for every t in $\mathbb{R}_+$

$$\{T_k \leq t\} = \{N_t \geq k\} \in \mathcal{F}_t$$

since N_t is in $\mathcal{F}_t$. Heuristically, T_k is a stopping time because it is possible to construct an alarm system that sounds exactly at the time of kth arrival. Another stopping time is

$$T = \inf\{\ t \geq a :\ N_t = N_{t-a}\ \},$$

where $a > 0$ is fixed, that is, the first time that an interval of length a passes without an arrival. We leave the proof to Exercise 1.35, because it needs tools to be developed below. Finally, here is a random time that is not a stopping time: fix $b > 0$, and let

$$L = \inf\{ \ t \in \mathbb{R}_+ : \ N_t = N_b \ \},$$

that is, L is the time of last arrival before the time b if there is one, and it is 0 if there is none.

Conventions for the end of time

Soon we shall introduce the concept of information accumulated by the time T, when the alarm sounds. Since stopping times can take $+\infty$ as a value, in case $+\infty$ is not in $\mathbb{T}$, we need to extend the definition of the filtration $\mathcal{F}$ on $\mathbb{T}$ onto $\bar{\mathbb{T}} = \mathbb{T} \cup \{+\infty\}$. We do so by letting $\mathcal{F}_\infty$, which we also denote by $\lim \mathcal{F}_t$, be defined as

1.8
$$\mathcal{F}_\infty = \lim \mathcal{F}_t = \bigvee_{t \in \mathbb{T}} \mathcal{F}_t,$$

the σ-algebra generated by the union of all the $\mathcal{F}_t$. Then, $(\mathcal{F}_t)_{t \in \bar{\mathbb{T}}}$ is a filtration on $\bar{\mathbb{T}}$, and T is a stopping time of it if and only if T is a stopping time of $(\mathcal{F}_t)_{t \in \mathbb{T}}$. Also, every adapted process X indexed by $\mathbb{T}$ can be extended onto $\bar{\mathbb{T}}$ by appending to X an arbitrary variable X_∞ picked from $\mathcal{F}_\infty$. We shall still write $\mathcal{F}$ for the extended filtration, especially since $\mathcal{F}_\infty$ has no information in it that was not in $(\mathcal{F}_t)_{t \in \mathbb{T}}$.

Past until T

Let $\mathcal{F}$ be a filtration on $\mathbb{T}$, extended to $\bar{\mathbb{T}}$ as above. Let T be a stopping time of it. Corresponding to the notion of the body of information accumulated by the time T, we define

1.9
$$\mathcal{F}_T = \{ \ H \in \mathcal{H} : \ H \cap \{T \le t\} \in \mathcal{F}_t \ \text{ for each } t \text{ in } \bar{\mathbb{T}} \ \}.$$

It is easy to check that $\mathcal{F}_T$ is a σ-algebra and that $\mathcal{F}_T \subset \mathcal{F}_\infty \subset \mathcal{H}$; it is called the *past until* T.

If T is a fixed time, say $T(\omega) = t$ for all ω for some constant t in $\bar{\mathbb{T}}$, then $\mathcal{F}_T = \mathcal{F}_t$; hence, there is no ambiguity in the notation $\mathcal{F}_T$.

For an arbitrary stopping time T, note that the event $\{T \le r\}$ belongs to $\mathcal{F}_T$ for every $r \ge 0$, because

$$\{T \le r\} \cap \{T \le t\} = \{T \le r \wedge t\} \in \mathcal{F}_t$$

for each t. Thus, T is $\mathcal{F}_T$-measurable.

As usual with σ-algebras, $\mathcal{F}_T$ will also denote the collection of all $\mathcal{F}_T$-measurable random variables. Heuristically, then, 1.9 is equivalent to saying that $\mathcal{F}_T$ consists of those $\bar{\mathbb{R}}$-valued variables V such that, for every possibility ω, the value $V(\omega)$ can be told by the time $T(\omega)$, the time of the alarm sound. The following is the precise version.

1.10 Theorem. *A random variable V belongs to $\mathcal{F}_T$ if and only if*

1.11
$$V\,1_{\{T\le t\}} \in \mathcal{F}_t$$

for every t in $\bar{\mathbb{T}}$. In particular, if $\bar{\mathbb{T}} = \bar{\mathbb{N}}$, the condition is equivalent to requiring that, for every n in $\bar{\mathbb{N}}$,

1.12
$$V \cdot 1_{\{T=n\}} \in \mathcal{F}_n.$$

Proof. We may and do assume that V is positive and let X_t be the random variable appearing in 1.11. Then, for all r in $\mathbb{R}_+$ and t in $\mathbb{T}$,

$$\{V > r\} \cap \{T \le t\} = \{X_t > r\}.$$

Thus, by the definition 1.9, the event $\{V > r\}$ is in $\mathcal{F}_T$ for all r if and only if the event $\{X_t > r\}$ is in $\mathcal{F}_t$ for all r for every t in $\bar{\mathbb{T}}$. That is, $V \in \mathcal{F}_T$ if and only if $X_t \in \mathcal{F}_t$ for every t in $\bar{\mathbb{T}}$, which is the claim of the first statement. The particular statement for the case $\bar{\mathbb{T}} = \bar{\mathbb{N}}$ is immediate upon noting that

$$V \cdot 1_{\{T=n\}} = \begin{cases} X_n - X_{n-1} & \text{if } n \in \mathbb{N} \\ X_\infty - \sum_{n\in\mathbb{N}}(X_n - X_{n-1}) & \text{if } n = +\infty. \end{cases}$$ □

Representation of $\mathcal{F}$ and $\mathcal{F}_T$

Let $\mathbb{T} \subset \bar{\mathbb{R}}$. Let $\mathcal{F}$ be a filtration on it, extended onto $\bar{\mathbb{T}} = \mathbb{T} \cup \{+\infty\}$ if $\mathbb{T}$ does not include the point $+\infty$. We identify $\mathcal{F}$ with the collection of all right-continuous processes on $\bar{\mathbb{T}}$ that are adapted to $\mathcal{F}$. More precisely, we say that $X \in \mathcal{F}$ if

1.13 a) $X = (X_t)_{t\in\bar{\mathbb{T}}}$ is adapted to $\mathcal{F} = (\mathcal{F}_t)_{t\in\bar{\mathbb{T}}}$, and
b) the path $t \mapsto X_t(\omega)$ from $\bar{\mathbb{T}}$ into $\bar{\mathbb{R}}$ is right-continuous for each ω in Ω.

Remark. If $\mathbb{T}$ is $\mathbb{N}$ or $\bar{\mathbb{N}}$, then the condition (b) above holds automatically, because every path $n \mapsto X_n(\omega)$ is continuous in the discrete topology on $\bar{\mathbb{N}}$, which is the topology induced on $\bar{\mathbb{N}}$ by the ordinary topology of $\bar{\mathbb{R}}_+$. Consequently, in these cases,

$$X \in \mathcal{F} \iff X_n \in \mathcal{F}_n \ \text{ for each } n \text{ in } \bar{\mathbb{N}},$$

and the notation $X \in \mathcal{F}$ is amply justified.

The following characterization theorem shows the economy of thought achieved by this device: $\mathcal{F}_T$ consists of the values X_T of processes X in $\mathcal{F}$ at the time T. For a much simpler proof in the case of discrete time, see Exercise 1.32.

1.14 THEOREM. *Let T be a stopping time of $\mathcal{F}$, Then,*

$$\mathcal{F}_T = \{\ X_T :\ \ X \in \mathcal{F}\ \}.$$

Proof. a) Let $V \in \mathcal{F}_T$. Define $X_t = V\ 1_{\{T \le t\}}$, $t \in \bar{\mathbb{T}}$. Then, X is adapted to $\mathcal{F}$ by Theorem 1.10 and is obviously right-continuous, that is, $X \in \mathcal{F}$. Clearly $X_T = V$. So, $\mathcal{F}_T \subset \{\ X_T :\ \ X \in \mathcal{F}\ \}$.

b) To show the converse that $\mathcal{F}_T \supset \{\ X_T :\ \ X \in \mathcal{F}\ \}$, we let $X \in \mathcal{F}$, put $V = X_T$, and proceed to show that $V \in \mathcal{F}_T$. To that end, in view of Theorem 1.10, it is enough to show that $V\ 1_{\{T \le t\}} \in \mathcal{F}_t$ for every t in $\bar{\mathbb{T}}$. Fix t, and note that this is equivalent to showing that the mapping

$$h:\ \omega \mapsto V(\omega)\ \text{ from } \Omega_t = \{T \le t\} \text{ into } \bar{\mathbb{R}}$$

is $\hat{\mathcal{F}}_t$-measurable, where $\hat{\mathcal{F}}_t$ is the trace of $\mathcal{F}_t$ on Ω_t.

Let $B_s = \bar{\mathbb{T}} \cap [0, s]$ for $s \le t$ and let $\mathcal{B}_t = \mathcal{B}(B_t)$. Let f be the mapping $\omega \mapsto (T(\omega), \omega)$ from Ω_t into $B_t \times \Omega$. If $s \in B_t$ and $H \in \mathcal{F}_t$, then the inverse image of the rectangle $B_s \times H$ under f is the event $\{T \le s\} \cap H$, which event is in $\mathcal{F}_t$. Thus f is measurable with respect to $\hat{\mathcal{F}}_t$ and $\mathcal{B}_t \otimes \mathcal{F}_t$.

Let g be the mapping $(s, w) \mapsto X_s(\omega)$ from $B_t \times \Omega$ into $\bar{\mathbb{R}}$. For each s, since X is adapted to $\mathcal{F}$, the mapping $\omega \mapsto X_s(\omega)$ is in $\mathcal{F}_s$ and, therefore, is in $\mathcal{F}_t$; and for each ω, by the way X is chosen, the mapping $s \mapsto X_s(\omega)$ is right-continuous on B_t. Thus, g is $\mathcal{B}_t \otimes \mathcal{F}_t$-measurable (see Exercise I.6.31 for this). It follows that the mapping $g \circ f$ from Ω_t into $\bar{\mathbb{R}}$ is $\hat{\mathcal{F}}_t$-measurable. But, $g \circ f(\omega) = g(T(\omega), \omega) = X_{T(\omega)}(\omega) = V(\omega) = h(\omega)$ for ω in Ω_t. Thus h is $\hat{\mathcal{F}}_t$-measurable as needed to complete the proof. □

1.15 REMARK. *Progressiveness.* The preceding theorem can be rephrased: $V \in \mathcal{F}_T$ if and only if $V = X_T$ for some right-continuous process X adapted to $\mathcal{F}$. This does not exclude the possibility that there is some other process Y, not right-continuous, such that $V = Y_T$ as well. Indeed, the last paragraph of the preceding proof shows what is required of Y: For each t, the mapping $(s, \omega) \mapsto Y_s(\omega)$ from $B_t \times \Omega$ into $\bar{\mathbb{R}}$ should be $\mathcal{B}_t \otimes \mathcal{F}_t$-measurable. Such processes Y are said to be $\mathcal{F}$-*progressive*. So, in fact, $V \in \mathcal{F}_T$ if and only if $V = Y_T$ for some $\mathcal{F}$-progressive process Y. Of course, every right-continuous adapted process is progressive. In discrete time, if $\mathbb{T}$ is discrete, every process is in fact continuous and, hence, every adapted process is progressive.

Comparing different pasts

If S and T are stopping times of $\mathcal{F}$, and if S is dominated by T (that is, $S(\omega) \le T(\omega)$ for all ω), then the information accumulated by the time

S should be less than that accumulated by T. The following shows this and gives further comparisons for general S and T.

1.16 THEOREM. *Let S and T be stopping times of $\mathcal{F}$. Then,*

a) *$S \wedge T$ and $S \vee T$ are stopping times of $\mathcal{F}$;*
b) *if $S \leq T$ then $\mathcal{F}_S \subset \mathcal{F}_T$;*
c) *in general, $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$; and*
d) *if $V \in \mathcal{F}_S$ then the following are in $\mathcal{F}_{S\wedge T}$:*

$$V\ 1_{\{S\leq T\}},\ V\ 1_{\{S=T\}},\ V\ 1_{\{S<T\}}.$$

Proof. i) Since S and T are stopping times, the events $\{S \leq t\}$ and $\{T \leq t\}$ are in $\mathcal{F}_t$ for every time t. Therefore, so are the events $\{S \wedge T \leq t\} = \{S \leq t\} \cup \{T \leq t\}$ and $\{S \vee T \leq t\} = \{S \leq t\} \cap \{T \leq t\}$. Hence, $S \wedge T$ and $S \vee T$ are stopping times. This proves (a).

ii) Let $V \in \mathcal{F}_S$. By Theorem 1.10,

1.17 $$X_t = V\ 1_{\{S\leq t\}}, \quad t \in \bar{\mathbb{T}},$$

defines a process X adapted to $\mathcal{F}$. Clearly, X is right-continuous. Hence, $X \in \mathcal{F}$ in the sense of 1.13.

iii) If $S \leq T$, then $X_T = V$ by 1.17, and $X_T \in \mathcal{F}_T$ by Theorem 1.14 since $X \in \mathcal{F}$. So, if $S \leq T$ then $\mathcal{F}_S \subset \mathcal{F}_T$. This proves (b).

iv) We prove (d) next. Let the stopping times S and T be arbitrary. Then $S\wedge T$ is a stopping time by part (a), and $X_{S\wedge T} \in \mathcal{F}_{S\wedge T}$ by Theorem 1.14. Hence, replacing t in 1.17 with $S \wedge T$ we see that

1.18 $$V\ 1_{\{S\leq T\}} \in \mathcal{F}_{S\wedge T}.$$

In particular, taking $V = 1$ in 1.18 shows that the event $\{S \leq T\}$ belongs to $\mathcal{F}_{S\wedge T}$. By symmetry, then, so does the event $\{T \leq S\}$. Hence, so do the events $\{S = T\} = \{S \leq T\} \cap \{T \leq S\}$ and $\{S < T\} = \{S \leq T\} \setminus \{S = T\}$. It follows that multiplying the left side of 1.18 with the indicator of $\{S = T\}$ or with the indicator of $\{S < T\}$ will not alter the membership in $\mathcal{F}_{S\wedge T}$. This proves (d).

v) There remains to prove (c) with S and T arbitrary. Since the stopping time $S \wedge T$ is dominated by both S and T, we have $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_S$ and $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_T$ by part (b) proved above. Hence $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_S \cap \mathcal{F}_T$. To prove the converse containment, let H be an event in $\mathcal{F}_S \cap \mathcal{F}_T$. Then, by part (d) proved above, $H \cap \{S \leq T\} \in \mathcal{F}_{S\wedge T}$ since H is in $\mathcal{F}_S$, and $H \cap \{T \leq S\} \in \mathcal{F}_{S\wedge T}$ since H is in $\mathcal{F}_T$, hence, their union, which is H, belongs to $\mathcal{F}_{S\wedge T}$. So, $\mathcal{F}_S \cap \mathcal{F}_T \subset \mathcal{F}_{S\wedge T}$. □

Times foretold

Let S be a stopping time of $\mathcal{F}$. Let T be a random time such that $T \geq S$ but whose value can be told by the time S, that is, $T \in \mathcal{F}_S$. Then, T is said to be *foretold* by S. Obviously, T is again a stopping time of $\mathcal{F}$. For example, if t is deterministic, then $S + t \in \mathcal{F}_S$ and $S + t \geq S$, so $S + t$ is foretold by S and is a stopping time.

Approximation by discrete stopping times

Discrete stopping times, that is, stopping times that take values in a countable subset of $\bar{\mathbb{R}}$, are generally easier to work with. The following constructs a sequence of such times that approximates a given stopping time with values in $\bar{\mathbb{R}}_+$.

We start by defining, for each integer n in $\mathbb{N}$,

1.19 $$d_n(t) = \begin{cases} \frac{k+1}{2^n} & \text{if } \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \text{ for some } k \text{ in } \mathbb{N}, \\ +\infty & \text{if } t = +\infty. \end{cases}$$

Then, $d_n : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ is a step function, it is increasing and right-continuous, and $d_n(t) > t$ for every $t < \infty$. Further, $d_0 \geq d_1 \geq d_2 \geq \cdots$ with $\lim d_n(t) = t$ for each t in $\bar{\mathbb{R}}_+$.

1.20 PROPOSITION. *Let $\mathcal{F}$ be a filtration on $\bar{\mathbb{R}}_+$ and let T be a stopping time of it. Define*

$$T_n = d_n \circ T, \quad n \in \mathbb{N}.$$

Then (T_n) is a sequence of discrete stopping times of $\mathcal{F}$ which decreases to T.

Proof. Fix n. Being a measurable function of T, the random time T_n belongs to $\mathcal{F}_T$. Since $d_n(t) > t$ for all $t < \infty$ and $d_n(\infty) = \infty$, we have $T_n \geq T$. Thus, T_n is foretold by T and is a stopping time of $\mathcal{F}$. Obviously, it is discrete. Since $d_n(t)$ decreases to t as $n \to \infty$, the sequence (T_n) decreases to T. □

Conditioning at stopping times

This refers to conditional expectations given $\mathcal{F}_T$, where T is a stopping time of the filtration $\mathcal{F}$. Since $\mathcal{F}_T$ represents the total information by the time T, we think of $\mathbb{E}_{\mathcal{F}_T} X = \mathbb{E}(X|\mathcal{F}_T)$ as our estimate of X at time T, based on the information available then. To indicate this point of view, and also to lighten the notation somewhat, we adopt the following notational device:

1.21 CONVENTION. *We write $\mathbb{E}_T$ for $\mathbb{E}_{\mathcal{F}_T} = \mathbb{E}(\cdot|\mathcal{F}_T)$.*

In particular, every deterministic time t is a stopping time, and the notation makes sense for such t as well: $\mathbb{E}_t$ is the short notation for $\mathbb{E}_{\mathcal{F}_t} = \mathbb{E}(\cdot|\mathcal{F}_t)$. The following is a summary, in this context and notation, of the definition and various properties of the conditional expectations given $\mathcal{F}_T$.

1.22 THEOREM. *The following hold for all positive random variables X, Y, Z and all stopping times S and T of $\mathcal{F}$:*

a) Defining property*: $\mathbb{E}_T X = Y$ if and only if $Y \in \mathcal{F}_T$ and $\mathbb{E}\ VX = \mathbb{E}\ VY$ for every positive V in $\mathcal{F}_T$.*

b) Unconditioning*: $\mathbb{E}\ \mathbb{E}_T X = \mathbb{E}\ X$.*

c) Repeated conditioning*: $\mathbb{E}_S \mathbb{E}_T X = \mathbb{E}_{S\wedge T} X$.*

d) Conditional determinism*: $\mathbb{E}_T(X + YZ) = X + Y\mathbb{E}_T Z$ if $X, Y \in \mathcal{F}_T$.*

REMARK. The positivity condition on X, Y, Z ensures that the conditional expectations are well-defined. The properties above can be extended to integrable X, Y, Z and further, once one makes sure that the conditional expectations involved do exist. Of course, in the defining property, V can be limited to indicators in $\mathcal{F}_T$.

Proof. Except for the claim on repeated conditioning, all these are no more than re-wordings of the definition of conditional expectations and Theorem IV.1.10.

To show the claim regarding repeated conditioning, we start with the following observation: If $S \le T$ then $\mathcal{F}_S \subset \mathcal{F}_T$ by Theorem 1.16 above, and Theorem IV.1.10 applies to show that $\mathbb{E}_S\mathbb{E}_T = \mathbb{E}_S$. For arbitrary stopping times S and T, the preceding observation applies with the stopping times $S \wedge T \le T$ to yield $\mathbb{E}_{S\wedge T}\mathbb{E}_T = \mathbb{E}_{S\wedge T}$. Thus, putting

1.23 $$Y = \mathbb{E}_T X,$$

we see that the claim to be proved reduces to showing that

1.24 $$\mathbb{E}_S Y = \mathbb{E}_{S\wedge T} Y.$$

The right side of 1.24 is a random variable in $\mathcal{F}_{S\wedge T}$, and $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_S$ since $S \wedge T \le S$; thus, the right side is in $\mathcal{F}_S$ and, hence, has the required measurability to be a candidate for $\mathbb{E}_S Y$. To complete the proof of 1.24, there remains to show that

1.25 $$\mathbb{E}\ VY = \mathbb{E}\ V\ \mathbb{E}_{S\wedge T} Y$$

for every positive V in $\mathcal{F}_S$ (see the defining property for $\mathbb{E}_S$).

Fix V such. Then, $V\ 1_{\{S\le T\}} \in \mathcal{F}_{S\wedge T}$ by Theorem 1.16d, and the defining property for $\mathbb{E}_{S\wedge T}$ yields

1.26 $$\mathbb{E}\ V\ 1_{\{S\le T\}} Y = \mathbb{E}\ V\ 1_{\{S\le T\}}\ \mathbb{E}_{S\wedge T} Y.$$

On the other hand, since $Y \in \mathcal{F}_T$ by its definition 1.23, Theorem 1.16d shows that $Y\ 1_{\{T<S\}} \in \mathcal{F}_{S\wedge T}$, and the conditional determinism yields

1.27 $$\mathbb{E}\ VY1_{\{T<S\}} = \mathbb{E}\ V\mathbb{E}_{S\wedge T} Y1_{\{T<S\}} = \mathbb{E}\ V1_{\{T<S\}}\mathbb{E}_{S\wedge T} Y.$$

Adding 1.26 and 1.27 side by side yields the desired equality 1.25. □

Exercises

1.28 *Galmarino's test.* Let X be a continuous stochastic process with index set $\mathbb{R}_+$ and state space $\mathbb{R}$. Let $\mathcal{F}$ be the filtration generated by X. Show that a random time T is a stopping time of $\mathcal{F}$ if and only if, for every pair of outcomes ω and ω',

$$T(\omega) = t, \quad X_s(\omega) = X_s(\omega') \text{ for all } s \leq t \Rightarrow T(\omega') = t.$$

1.29 *Entrance times.* Let X and $\mathcal{F}$ be as in 1.28. For fixed $b \geq 0$, let T be the time of first entrance to $[b, \infty]$, that is,

$$T = \inf\{t \in \mathbb{R}_+ : \ X_t \geq b\}.$$

Show that T is a stopping time of $\mathcal{F}$. Show that, in general,

$$T = \inf\{t \in \mathbb{R}_+ : \ X_t > b\}$$

is *not* a stopping time of $\mathcal{F}$.

1.30 *Past until T.* Show that $\mathcal{F}_T$ defined by 1.9 is indeed a σ-algebra on Ω.

1.31 *Strict past at T.* Let T be a stopping time of $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$. Let $\hat{\mathcal{F}}_t$ be the trace of $\mathcal{F}_t$ on $\{t < T\}$, that is, $\hat{\mathcal{F}}_t$ consists of events of the form $H \cap \{t < T\}$ with H in $\mathcal{F}_t$. Let $\mathcal{F}_{T-}$ be the σ-algebra generated by $\cup_t \hat{\mathcal{F}}_t$. Unlike $\mathcal{F}_T$, events in $\mathcal{F}_{T-}$ do not have explicit representations. Show that $\mathcal{F}_{T-} \subset \mathcal{F}_T$.

1.32 *Characterization of $\mathcal{F}_T$ in discrete time.* Prove Theorem 1.14 directly when $\mathbb{T} = \mathbb{N}$. Hints: $V \in \mathcal{F}_T \iff V 1_{\{T=n\}} \in \mathcal{F}_n$ for every n in $\mathbb{N}$; and if $X \in \mathcal{F}$ then $X_T 1_{\{T=n\}} = X_n 1_{\{T=n\}}$.

1.33 *Stopping times foretold.* Let S and T be stopping times. Show that $S+T$ is foretold by $S \vee T$ and thus is a stopping time.

1.34 *Supremums.* Let T_n be a stopping time for each n in $\mathbb{N}^*$. Show that, then, $\sup T_n$ is again a stopping time of $\mathcal{F}$. A similar claim for $\inf T_n$ is generally false; see, however, Proposition 7.9.

1.35 *Arrival processes.* In Example 1.7, observe that, for every $t < \infty$, we have $T_k(\omega) \leq t < T_{k+1}(\omega)$ for some integer k depending on t and ω. Recall that every T_k is a stopping time of $\mathcal{F}$, the filtration generated by N. Put $T_0 = 0$ for convenience. Let T be as defined in 1.7. Note that, for every ω, $T(\omega) = T_k(\omega) + a$ for some k.

a) Show that, for each k in $\mathbb{N}$,

$$\{T = T_k + a\} = \{T_1 - T_0 \leq a, \ldots, T_k - T_{k-1} \leq a\} \cap \{T_{k+1} > T_k + a\}.$$

Show that this event is in $\mathcal{F}_{T_k + a}$. Conclude that, for every t in $\mathbb{R}_+$,

$$\{T = T_k + a\} \cap \{T \leq t\} \in \mathcal{F}_t.$$

b) Show that T is a stopping time of $\mathcal{F}$.

1.36 *Continuation.* Now we regard $0 = T_0 < T_1 < T_2 < \cdots$ as the times of successive replacements for some device. Then N_t becomes the number of replacements during $(0, t]$, and we define A_t to be the age of the unit in use at time t:

$$A_t(\omega) = t - T_k(\omega) \qquad \text{if } T_k(\omega) \le t < T_{k+1}(\omega).$$

a) Show that $t \mapsto A_t(\omega)$ is strictly increasing and continuous everywhere on $\mathbb{R}_+$ except for downward jumps to 0 at times $T_1(\omega), T_2(\omega), \ldots$. At these times, it is right-continuous.

b) Show that the process $A = (A_t)_{t \in \mathbb{R}_+}$ is adapted to $\mathcal{F}$.

c) Show that $T = \inf\{t \in \mathbb{R}_+ : \; A_t \ge a\}$, and show that T is a stopping time of $\mathcal{F}$ by a direct reasoning using this relationship to A.

2 Martingales

Martingales are the mainstay and unifying force underlying much of the theory of stochastic processes. This section is to introduce them and give some examples from Markov chains, Brownian motion, and Poisson processes.

Let $\mathbb{T}$ be a subset of $\bar{\mathbb{R}}$, let $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ be a filtration over $\mathbb{T}$ extended onto $\bar{\mathbb{T}} = \mathbb{T} \cup \{+\infty\}$ by 1.8 if $+\infty$ is not in $\mathbb{T}$, and recall the notational convention 1.21 regarding conditional expectations given $\mathcal{F}_T$.

2.1 Definition. *A real-valued stochastic process $X = (X_t)_{t \in \mathbb{T}}$ is called an $\mathcal{F}$-*submartingale *if X is adapted to $\mathcal{F}$, each X_t is integrable, and*

$$\mathbb{E}_s(X_t - X_s) \ge 0 \tag{2.2}$$

*whenever $s < t$. It is called an $\mathcal{F}$-*supermartingale *if $-X$ is an $\mathcal{F}$-submartingale, and an $\mathcal{F}$-martingale if it is both an $\mathcal{F}$-submartingale and an $\mathcal{F}$-supermartingale.*

Adaptedness and integrability are regularity conditions; they remain the same for submartingales, supermartingales, and martingales. The essential condition is 2.2: Given the information $\mathcal{F}_s$, the conditional expectation of the future increment $X_t - X_s$ is positive for submartingales, negative for supermartingales, and zero for martingales.

Indeed, since $X_s \in \mathcal{F}_s$, the conditional determinism property yields $\mathbb{E}_s X_s = X_s$, which shows that the parentheses around $X_t - X_s$ are superfluous; they are put there to make us think in terms of the increments. So, 2.2 can be re-written as

$$\mathbb{E}_s X_t \ge X_s, \qquad s < t\;; \tag{2.3}$$

this is for submartingales. The inequality is reversed for supermartingales and becomes an equality for martingales. Thus, roughly speaking, submartingales have a systematic tendency to be increasing, supermartingales to be decreasing, and martingales to be neither increasing nor decreasing. See Theorem 3.2 below for a sharper version of this remark.

2.4 Remarks. a) Let X be an $\mathcal{F}$-submartingale. For $s < t < u$ in $\mathbb{T}$,

$$\mathbb{E}_s(X_u - X_t) = \mathbb{E}_s\mathbb{E}_t(X_u - X_t) \geq \mathbb{E}_s 0 = 0$$

by 1.22c on repeated conditioning and the submartingale inequality $\mathbb{E}_t(X_u - X_t) \geq 0$. That is, given the cumulative information $\mathcal{F}_s$ available at the present time s, the estimate of any remote future increment is positive. Obviously, if X is a martingale, the inequality becomes an equality.

b) When the index set $\mathbb{T}$ is discrete, the reasoning of the preceding remark shows that it is sufficient to check the inequality 2.2 for times s and t that are next to each other, and then 2.2 holds for arbitrary $t > s$. For instance, when $\mathbb{T} = \mathbb{N}$, the martingale equality $\mathbb{E}_s(X_t - X_s) = 0$ holds if and only if

$$\mathbb{E}_n(X_{n+1} - X_n) = 0, \qquad n \in \mathbb{N}.$$

c) Let X be an $\mathcal{F}$-submartingale. For $s < t$, the random variable $\mathbb{E}_s(X_t - X_s)$ is positive and, therefore, is almost surely zero if and only if its expectation $\mathbb{E}\ \mathbb{E}_s(X_t - X_s)$ is zero. Since $\mathbb{E}\ \mathbb{E}_s = \mathbb{E}$, it follows that the submartingale X is in fact a martingale if $\mathbb{E}X_t = \mathbb{E}X_0$ for all times t.

d) If X and Y are $\mathcal{F}$-submartingales, then so is $aX + bY$ for a and b in $\mathbb{R}_+$. If X and Y are martingales, then so is $aX + bY$ for a and b in $\mathbb{R}$.

e) If X and Y are $\mathcal{F}$-submartingales, then so is $X \vee Y$, where $X \vee Y = (X_t \vee Y_t)_{t \in \mathbb{T}}$. If X and Y are $\mathcal{F}$-supermartingales, then so is $X \wedge Y$.

f) Let f be a convex function on $\mathbb{R}$. If X is an $\mathcal{F}$-martingale and if $f \circ X_t$ is integrable for every time t, then $f \circ X$ is an $\mathcal{F}$-submartingale. This follows from Jensen's inequality for conditional expectations (see IV.1.8): for $s < t$,

$$\mathbb{E}_s\ f \circ X_t \geq f \circ (\mathbb{E}_s X_t) = f \circ X_s$$

since $\mathbb{E}_s X_t = X_s$ for martingales. In particular, if X is a martingale, then $X^+ = (X_t^+)$ and $X^- = (X_t^-)$ and $|X| = (|X_t|)$ are submartingales, and so is $|X|^p = (|X_t|^p)$ provided that $\mathbb{E}|X_t|^p < \infty$ for every time t.

g) Similarly, if f is convex and increasing, and if X is an $\mathcal{F}$-submartingale with $f \circ X_t$ integrable for all t, then $f \circ X$ is again an $\mathcal{F}$-submartingale. In particular, if X is a submartingale, so is X^+.

h) Since $\mathbb{E}_s(X_t - X_s)$ belongs to $\mathcal{F}_s$, it is positive if and only if its integral over every event H in $\mathcal{F}_s$ is positive. Thus, the submartingale inequality 2.2 is equivalent to the following:

$$\mathbb{E}(X_t - X_s)\ 1_H \geq 0, \qquad H \in \mathcal{F}_s, s < t.$$

i) Let X be an $\mathcal{F}$-submartingale. Let $\mathcal{G}$ be the filtration generated by X. Then, X is automatically adapted to $\mathcal{G}$ and is integrable, and

$$\mathbb{E}_{\mathcal{G}_s}(X_t - X_s) = \mathbb{E}_{\mathcal{G}_s}\mathbb{E}_{\mathcal{F}_s}(X_t - X_s) = \mathbb{E}_{\mathcal{G}_s}\mathbb{E}_s(X_t - X_s) \geq 0$$

by the repeated conditioning property since $\mathcal{G}_s \subset \mathcal{F}_s$. Thus, X is a $\mathcal{G}$-submartingale.

Examples of martingales

2.5 *Sums of independent variables.* Let $X_1, X_2, \ldots$ be independent random variables with mean 0. Let $S_0 = 0$ and put $S_n = S_0 + X_1 + \cdots + X_n$ for $n \geq 1$. Let $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be the filtration generated by $S = (S_n)_{n\in\mathbb{N}}$. Then S is adapted to $\mathcal{F}$ trivially, and each S_n is integrable (with mean 0), and

$$\mathbb{E}_n(S_{n+1} - S_n) = \mathbb{E}_n X_{n+1} = \mathbb{E}\ X_{n+1} = 0,$$

since X_{n+1} is independent of $\mathcal{F}_n$ and has mean 0. Thus, S is a martingale; see Remark 2.4b. Much of classical martingale theory is an extension of this case.

2.6 *Products of independent variables.* Let $R_1, R_2, \ldots$ be independent random variables with mean 1 and some finite variance. Let $M_0 = 1$ and

$$M_n = M_0 R_1 R_2 \cdots R_n, \qquad n \in \mathbb{N}.$$

Let $\mathcal{F}$ be the filtration generated by $M = (M_n)_{n\in\mathbb{N}}$. Then, M is adapted to $\mathcal{F}$ trivially, and each M_n is integrable in view of Schwartz's inequality (see Theorem II.3.6a) and the assumption that the R_n have finite variances. Also,

$$\mathbb{E}_n M_{n+1} = \mathbb{E}_n M_n R_{n+1} = M_n \mathbb{E}_n R_{n+1} = M_n$$

by the independence of R_{n+1} from $\mathcal{F}_n$ and the hypothesis that $\mathbb{E}\ R_{n+1} = 1$. Hence, M is an $\mathcal{F}$-martingale via Remarks 2.3 and 2.4b.

In the further case where the R_n are positive, the martingale M is considered to be a reasonable model for the evolution of the price of a share of stock. Then, M_n stands for the price of a share at time n, and R_{n+1} is interpreted as the return at time $n+1$ per dollar invested at time n in that stock. The economists' argument for the martingale equality is as follows (very roughly): The information $\mathcal{F}_n$ is available to the whole market. If the conditional expectation $\mathbb{E}_n(M_{n+1} - M_n)$ were strictly positive over some event H, then there would have been a rush to buy which would have forced M_n to go higher; if the expectation were strictly negative over some event, then there would have been a rush to sell and M_n would go lower; the equilibrium attains only if the conditional expectation is zero over an almost sure set.

Uniformly integrable martingales

These are martingales that are also uniformly integrable. They play the central role in martingale theory. The next proposition shows how to obtain such a martingale: take an integrable random variable and let X_t be our estimate of it at time t, given the information $\mathcal{F}_t$ accumulated until then. Conversely, it will be shown in Theorems 4.7 and 5.13 that every uniformly integrable martingale is obtained in this manner. Here, $\mathbb{T} \subset \bar{\mathbb{R}}$ is arbitrary and $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is a filtration on $\mathbb{T}$; see Definition II.3.12 *et seq.* for uniform integrability.

2.7 Proposition. *Let Z be an integrable random variable. Define*

$$X_t = \mathbb{E}_t Z, \qquad t \in \mathbb{T}.$$

Then $X = (X_t)_{t\in\mathbb{T}}$ is an $\mathcal{F}$-martingale and is uniformly integrable.

Proof. Adaptedness is immediate, since $\mathbb{E}_t Z$ is in $\mathcal{F}_t$ by the definition of conditional expectations. Each X_t is integrable, because Z is so and the conditional expectation of an integrable variable is integrable. The martingale equality follows from the properties of repeated conditioning: for times $s < t$,

$$\mathbb{E}_s X_t = \mathbb{E}_s \mathbb{E}_t Z = \mathbb{E}_s Z = X_s.$$

Finally, the uniform integrability of the collection (X_t) follows from the following more general result of independent interest. □

2.8 Lemma. *Let Z be an integrable random variable. Then,*

$$\mathcal{K} = \{X : X = \mathbb{E}_\mathcal{G} Z \text{ for some sub-}\sigma\text{-algebra } \mathcal{G} \text{ of } \mathcal{H}\}$$

is uniformly integrable.

Proof. Since Z is integrable, the singleton $\{Z\}$ is uniformly integrable. Thus, by Theorem II.3.19, there is an increasing convex function f with $\lim_{x\to\infty} f(x)/x = +\infty$ such that $\mathbb{E}\, f\circ|Z| < \infty$. We show next that, with the same f,

2.9
$$\mathbb{E}\, f\circ|X| \le \mathbb{E}\, f\circ|Z|$$

for every X in $\mathcal{K}$, which implies, via Theorem II.3.19 again, that $\mathcal{K}$ is uniformly integrable.

Let $X = \mathbb{E}_\mathcal{G} Z$ for some sub-σ-algebra $\mathcal{G}$ of $\mathcal{H}$. Then, by Jensen's inequality IV.1.8,

$$|X| = |\mathbb{E}_\mathcal{G} Z| \le \mathbb{E}_\mathcal{G}|Z|.$$

Thus, since f is increasing and convex,

$$f\circ|X| \le f\circ(\mathbb{E}_\mathcal{G}|Z|) \le \mathbb{E}_\mathcal{G} f\circ|Z|,$$

where the last inequality is Jensen's again. Now, taking expectations on both sides and recalling that $\mathbb{E}\,\mathbb{E}_\mathcal{G} = \mathbb{E}$, we obtain the desired end 2.9. □

Markov chains

Here, the index set is $\mathbb{N}$, and $\mathcal{F}$ is a filtration over $\mathbb{N}$. Let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process with state space $(E, \mathcal{E})$, and let P be a Markov kernel on $(E, \mathcal{E})$; see section I.6 for the latter and recall the notation

2.10
$$Pf(x) = \int_E P(x, dy)\, f(y), \qquad x \in E, f \in \mathcal{E}_+.$$

2.11 DEFINITION. *The process X is called a Markov chain with transition kernel P, with respect to $\mathcal{F}$, if X is adapted to $\mathcal{F}$ and*

2.12 $$\mathbb{E}_n \, f \circ X_{n+1} = (Pf) \circ X_n$$

for every function f in $\mathcal{E}_+$ and time n in $\mathbb{N}$.

Markov chains have an extensive theory; see also Chapter IV, Section 5, for many variations. Much of their theory (and the theory of their continuous-time counterparts) has been influenced strongly by its connections to classical potential theory. As a result, harmonic and subharmonic and superharmonic functions of the classical theory have found their counterparts for Markov processes and, through Markov processes, have influenced the definitions of martingales and submartingales and supermartingales. Here is the connection.

Let X be a Markov chain, with respect to some filtration $\mathcal{F}$, with state space $(E, \mathcal{E})$ and transition kernel P. A bounded function f in $\mathcal{E}$ is said to be *harmonic*, *subharmonic*, and *superharmonic* if, respectively,

2.13 $$f = Pf, \quad f \le Pf, \quad f \ge Pf.$$

Put $M_n = f \circ X_n$; it is integrable since f is bounded, and it is obviously in $\mathcal{F}_n$. Indeed, $M = (M_n)_{n \in \mathbb{N}}$ is a martingale if f is harmonic, a submartingale if f is subharmonic, and a supermartingale if f is superharmonic. Here is the proof of the supermartingale inequality assuming that f is superharmonic; the other two cases can be shown similarly.

$$\mathbb{E}_n M_{n+1} = \mathbb{E}_n f \circ X_{n+1} = (Pf) \circ X_n \le f \circ X_n = M_n,$$

where we used the Markov property 2.12 to justify the second equality and the superharmonicity ($Pf \le f$) to justify the inequality.

A more recent connection is the following characterization of Markov chains in terms of martingales; this becomes a deep result in continuous-time.

2.14 THEOREM. *Let X be adapted to $\mathcal{F}$. Then X is a Markov chain with transition kernel P with respect to $\mathcal{F}$ if and only if*

$$M_n = f \circ X_n - \sum_{m=0}^{n-1} (Pf - f) \circ X_m, \qquad n \in \mathbb{N},$$

is a martingale with respect to $\mathcal{F}$ for every bounded f in $\mathcal{E}_+$.

Proof. Note that

$$M_{n+1} - M_n = f \circ X_{n+1} - (Pf) \circ X_n;$$

thus, $\mathbb{E}_n M_{n+1} - M_n = 0$ if and only if X has the Markov property 2.12. □

Wiener Processes

Let $\mathcal{F}$ be a filtration over $\mathbb{R}_+$. Let $W = (W_t)_{t \in \mathbb{R}_+}$ be a continuous process with state space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and starting point $W_0 = 0$.

2.15 Definition. *The continuous process W is called a Wiener process with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$ and*

2.16
$$\mathbb{E}_s\ f(W_{s+t} - W_s) = \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} f(x)$$

for all s and t in $\mathbb{R}_+$ and all positive Borel functions f on $\mathbb{R}$.

The defining relation 2.16 has three statements in it: the increment $W_{s+t} - W_s$ over the interval $(s, s+t]$ is independent of the past $\mathcal{F}_s$, the distribution of that increment is free of s, and the distribution is Gaussian with mean 0 and variance t. Indeed, 2.16 defines the probability law of W uniquely: for $0 = t_0 < t_1 < \cdots < t_n$, the probability law of $(W_{t_1}, \ldots, W_{t_n})$ is determined uniquely by the probability law of $(W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}})$, and the latter is the product of the distributions of $W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ by the independence of the increments, and the distributions are further specified by 2.16 as Gaussian with mean 0 and respective variances $t_1 - t_0, \ldots, t_n - t_{n-1}$. Incidentally, we see that W has stationary and independent increments (stationarity refers to the invariance of the distribution of $W_{s+t} - W_s$ as s varies). We shall study Wiener processes in Chapter VIII. Our aim at present is to introduce three martingales related to W. First is a useful characterization.

2.17 Proposition. *The process W is a Wiener process with respect to $\mathcal{F}$ if and only if, for each r in $\mathbb{R}$,*

$$M_t = \exp(rW_t - \tfrac{1}{2}r^2 t), \qquad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale.

Proof. *Necessity.* Suppose that W is Wiener. Then, a direct computation using 2.16 shows that, for $s < t$,

2.18
$$\mathbb{E}_s(M_t/M_s) = \mathbb{E}_s \exp[r(W_t - W_s) - \tfrac{1}{2}r^2(t-s)] = 1.$$

Thus, $\mathbb{E}_s M_t = M_s \mathbb{E}_s(M_t/M_s) = M_s$, which shows that M is a martingale (adaptedness and integrability being obvious).

Sufficiency. If M is a martingale, then $\mathbb{E}_s(M_t/M_s) = 1$, which means that 2.18 holds, or equivalently,

$$\mathbb{E}_s\ \exp\ r(W_{s+t} - W_s) = \exp\ \tfrac{1}{2}r^2 t.$$

This, being true for all r in $\mathbb{R}$, is equivalent to 2.16. □

It is worth noting that, for fixed r in $\mathbb{R}$, the process M of the preceding theorem is a continuous-time version of Example 2.6. Indeed, $M_{n+1} = M_n R_{n+1}$ where the random variable R_{n+1} now has a very specific distribution, namely, the distribution of $\exp[r(W_{n+1} - W_n - \frac{1}{2}r^2]$. Thus, the *exponential martingale* M is much used as a model for the evolution of stock prices. It is also the primary tool for studying Brownian motions by using results from martingale theory; see 5.20 *et seq.*

The next theorem gives the *martingale characterization of Wiener process.* We are able to prove here only the easy part, the necessity. For a proof of the sufficiency, see 6.21 to come.

2.19 THEOREM. *The continuous process W is Wiener with respect to $\mathcal{F}$ if and only if*

a) W *is an* $\mathcal{F}$*-martingale, and*
b) $Y = (W_t^2 - t)_{t\in\mathbb{R}_+}$ *is an* $\mathcal{F}$*-martingale.*

Proof of necessity. Let W be Wiener. Then, adaptedness and integrability conditions are obvious for W and Y. Now, the martingale equality for W is straightforward: for $s < t$, the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ and has mean 0; thus,

$$\mathbb{E}_s\ (W_t - W_s) = \mathbb{E}\ (W_t - W_s) = 0.$$

To show the martingale equality for Y, we first note that

$$Y_t - Y_s = (W_t - W_s)^2 + 2W_s(W_t - W_s) - (t - s)$$

and then use the facts that $W_s \in \mathcal{F}_s$ and that $W_t - W_s$ is independent of $\mathcal{F}_s$ and has mean 0 and variance $t - s$. Thus, as needed,

$$\mathbb{E}_s\ (Y_t - Y_s) = \mathbb{E}\ (W_t - W_s)^2 + 2W_s\ \mathbb{E}(W_t - W_s) - (t - s) = 0. \qquad \square$$

The Wiener process is the continuous martingale par excellence. It plays the same role in stochastic analysis as the Lebesgue measure does in ordinary analysis. In particular, every continuous martingale (in continuous-time) is obtained from a Wiener process by a random time change, just as most measures on $\mathbb{R}$ are obtained from the Lebesgue measure by a time change (see Theorem I.5.4).

Poisson martingales

Saying that a process is a martingale amounts to stating a property of it without specifying its probability law. However, on rare occasions, martingale property specifies the probability law. We stated, without proof, one such case: if W is a continuous martingale and if $W^2 - t$ is a martingale, then W is a Wiener process. Here, we provide another such case, even sharper, this time a pure-jump process.

Here, the index set is $\mathbb{R}_+$, and $\mathcal{F}$ is a filtration over it. Let $N = (N_t)_{t\in\mathbb{R}_+}$ be a counting process: this is a process with state space $(\mathbb{N}, 2^{\mathbb{N}})$ whose every path $t \mapsto N_t(\omega)$ starts from $N_0(\omega) = 0$, is increasing and right-continuous, and increases by jumps of size one only. Therefore, $N_t(\omega)$ is equal to the number of jumps of $s \mapsto N_s(\omega)$ in the interval $(0, t]$; Example 1.7 provides the complete picture. The following definition parallels that of the Wiener processes, Definition 2.15.

2.20 Definition. *The counting process N is said to be a Poisson process with rate c with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$ and*

2.21
$$\mathbb{E}_s\ f(N_{s+t} - N_s) = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} f(k)$$

for all s and t in $\mathbb{R}_+$ and all positive functions f on $\mathbb{N}$.

The defining equation 2.21 is equivalent to saying that the increment $N_{s+t} - N_s$ is independent of $\mathcal{F}_s$ and has the Poisson distribution with mean ct. As with Wiener processes, then, N has stationary and independent increments, and its probability law is completely determined by the positive constant c. Just as W is a martingale, for the Poisson process N, we have that $M = (N_t - ct)_{t\in\mathbb{R}_+}$ is an $\mathcal{F}$-martingale; this is immediate from 2.21:

2.22
$$\mathbb{E}_s(N_{s+t} - N_s) = ct, \qquad s, t \in \mathbb{R}_+.$$

It is surprising that, as the next theorem states, the simple property 2.22 is equivalent to 2.21. This is the *martingale characterization* theorem for Poisson processes; it parallels Theorem 2.19 and is even sharper.

2.23 Theorem. *Let N be a counting process. It is a Poisson process with rate c, with respect to $\mathcal{F}$, if and only if*

$$M_t = N_t - ct, \qquad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale.

The proof will be given in Section 6; see Proposition 6.13 and its proof.

Exercises and complements

2.24 *Restrictions.* Let $\mathbb{T}_0 \subset \bar{\mathbb{R}}$, and let $(X_t)_{t\in\mathbb{T}_0}$ be a martingale with respect to a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}_0}$. Then, for every $\mathbb{T}_1 \subset \mathbb{T}_0$, the process $(X_t)_{t\in\mathbb{T}_1}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}_1}$. The word "martingale" can be replaced with "submartingale" or with "supermartingale".

2.25 *Markov chains.* Let $X = (X_n)$ be a Markov chain with state space $(E, \mathcal{E})$ and transition kernel P. With P^n denoting the nth power of P–see I.6.6 for the definition–show that

$$\mathbb{E}_m\ f\circ X_{m+n} = (P^n f)\circ X_m, \qquad m, n \in \mathbb{N},$$

for every f in $\mathcal{E}_+$. Show that, for each fixed integer $k \geq 1$, $M_n = (P^{k-n}f)\circ X_n$ defines a martingale on $\mathbb{T} = \{0, 1, \ldots, k\}$.

2.26 *Poisson processes.* Let $N = (N_t)_{t \in \mathbb{R}_+}$ be a counting process adapted to some filtration $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$. Prove the following characterization theorem (see Proposition 2.17 for the parallel fo Wiener processes): N is a Poisson process with rate c with respect to $\mathcal{F}$ if and only if

$$M_t = \exp\ (-rN_t + ct - cte^{-r}), \qquad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale for every r in $\mathbb{R}_+$.

2.27 *Averages.* Let (X_n) be adapted to some filtration $(\mathcal{F}_n)$ and suppose that each X_n is integrable. Define

$$\bar{X}_n = \frac{1}{n+1}(X_0 + \cdots + X_n), \quad n \in \mathbb{N},$$

and assume that $\mathbb{E}_n X_{n+1} = \bar{X}_n$ for all n. Show that $(\bar{X}_n)$ is an $\mathcal{F}$-martingale.

2.28 *Positive supermartingales.* Let (X_n) be a positive supermartingale with respect to some filtration $(\mathcal{F}_n)$. Then, the following holds for almost every ω: if $X_m(\omega) = 0$ for some m, then $X_n(\omega) = 0$ for all $n \geq m$. Show this. Hint: Let $H = \{X_m = 0\}$ and show that $\mathbb{E}_m\ 1_H X_n = 0$ for $n \geq m$.

2.29 *Uniform integrability.* Let Z be an integrable random variable. Let $\mathcal{F} = (\mathcal{F}_t)_{t \in \bar{\mathbb{R}}_+}$ be a filtration on $\bar{\mathbb{R}}_+$. For each stopping time T of $\mathcal{F}$, let

$$X_T = \mathbb{E}_T\ Z,$$

Show that the collection $\{X_T :\ T \text{ is a stopping time of } \mathcal{F}\}$ is uniformly integrable.

2.30 *Martingales in L^p.* For p in $[1, \infty]$, a process X is said to be a martingale in L^p if, in addition to adaptedness and the martingale equality, the integrability condition for X_t is strengthened to requiring that $X_t \in L^p$ for every time t. Note that martingales in L^1 are simply martingales. Submartingales and supermartingales in L^p are defined similarly by replacing the condition $X_t \in L^1$ with the stronger condition that $X_t \in L^p$.

2.31 *L^p-boundedness.* A process (X_t) is said to be L^p-bounded if

$$\sup_t \mathbb{E}\ |X_t|^p < \infty.$$

With the notation $\|\cdot\|_p$ for the L^p-norm, the condition means that $\|X_t\|_p \leq c$ for some constant $c < \infty$. Recall: uniform integrability implies L^1-boundedness; the converse is generally false; but L^p-boundedness for some $p > 1$ implies uniform integrability.

2.32 *Square integrable martingales.* These are martingales that are L^2-bounded. This somewhat misleading usage seems well established.

2.33 *Quadratic variation.* Let (M_n) be a martingale in L^2 adapted to some filtration $(\mathcal{F}_n)$. Define an increasing process by setting $Q_0 = 0$ and

$$Q_{n+1} - Q_n = (M_{n+1} - M_n)^2, \qquad n \geq 0.$$

Show that the process X defined by

$$M_n^2 = M_0^2 + X_n + Q_n, \qquad n \geq 0,$$

is a martingale with $X_0 = 0$. Show that

$$\mathrm{Var} M_n = \mathrm{Var} M_0 + \mathbb{E} Q_n = \mathrm{Var} M_0 + \mathrm{Var}(M_1 - M_0) + \cdots + \mathrm{Var}(M_n - M_{n-1}).$$

The process Q is called the quadratic variation process for M.

3 Martingale Transformations and Maxima

This section contains the basic results for martingales in discrete time: integration in discrete time, Doob's stopping theorem, and inequalities for upcrossings and maxima. The index set is $\mathbb{N}$ unless stated otherwise; $\mathcal{F}$ is a filtration which we keep in the background; Convention 1.21 regarding conditional expectations is in force throughout; and all martingales, stopping times, and so on are with respect to the filtration $\mathcal{F}$.

Doob's decomposition

The object is to write a given process as the sum of a martingale and a predictable process, the latter to be defined presently. In the case of a submartingale, its predictable part turns out to be increasing, which clarifies our earlier statement that submartingales have a systematic tendency to be increasing.

3.1 Definition. *A process $F = (F_n)_{n\in\mathbb{N}}$ is said to be $\mathcal{F}$-predictable if $F_0 \in \mathcal{F}_0$ and $F_{n+1} \in \mathcal{F}_n$ for every n in $\mathbb{N}$.*

Heuristically, then, the cumulative information $\mathcal{F}_n$ available at time n determines the next value F_{n+1}, and thus, F is predictable in this dynamic sense. Note that every predictable process is adapted and more. The following is Doob's decomposition.

3.2 Theorem. *Let X be adapted and integrable. Then, it can be decomposed as*

3.3 $$X_n = X_0 + M_n + A_n, \qquad n \in \mathbb{N},$$

where M is a martingale with $M_0 = 0$, and A is predictable with $A_0 = 0$. This decomposition is unique up to equivalence. In particular, A is increasing if X is a submartingale, and decreasing if X is a supermartingale.

Proof. a) Put $M_0 = A_0 = 0$ and define M and A through their increments:

$$A_{n+1} - A_n = \mathbb{E}_n\,(X_{n+1} - X_n), \qquad M_{n+1} - M_n = (X_{n+1} - X_n) - (A_{n+1} - A_n)$$

for each $n \in \mathbb{N}$. Then, 3.3 holds, M is obviously a martingale, and A is predictable by the $\mathcal{F}_n$-measurability of the conditional expectation $\mathbb{E}_n\,(X_{n+1} - X_n)$. This proves the first statement.

b) If X is a submartingale, then 2.2 shows that $A_{n+1} - A_n \geq 0$, that is, A is increasing. If X is a supermartingale, then the inequality is reversed, and A is decreasing.

c) There remains to show the statement on uniqueness. To that end, let $X = X_0 + M' + A'$ be another such decomposition. Then $B = A - A' = M' - M$ is both predictable and a martingale. Thus,

$$B_{n+1} - B_n = \mathbb{E}_n\,(B_{n+1} - B_n) = 0, \qquad n \in \mathbb{N}\ ;$$

in other words, almost surely, $B_n = B_0 = 0$. Hence, almost surely, $A = A'$ and $M = M'$, as claimed. □

In Doob's decomposition, we have $X_{n+1} - X_n = A_{n+1} - A_n + M_{n+1} - M_n$; of these, $A_{n+1} - A_n$ is known by the time n; thus, the extra information gained by observing $X_{n+1} - X_n$ consists of the martingale increment $M_{n+1} - M_n$. For this reason, in engineering literature, A is called the prediction process, and M the *innovation* process.

Integration in discrete time

This is a resume of stochastic integration in discrete time. Let $M = (M_n)$ and $F = (F_n)$ be real-valued stochastic processes and define

3.4 $$X_n = M_0 F_0 + (M_1 - M_0)F_1 + \cdots + (M_n - M_{n-1})F_n, \qquad n \in \mathbb{N}.$$

Then, $X = (X_n)$ is called the integral of F with respect to M, or the *transform* of M by F, and we shall write

3.5 $$X = \int F\, dM$$

to indicate it. Indeed, F is a random function on $\mathbb{N}$, and M defines a random signed-measure on $\mathbb{N}$ which puts the mass $M_n - M_{n-1}$ at n except that the mass is M_0 at $n = 0$; then 3.4 is equivalent to writing

$$X_n = \int_{[0,n]} F\, dM,$$

the Lebesgue-Stieltjes integral of F over $[0, n]$ with respect to M. Such integrals are harder to define in continuous time, because Lebesgue-Stieltjes integrals make sense for M that are of bounded variation over bounded intervals, whereas most continuous martingales (including Wiener processes) have infinite variation over every open interval. Here we are working with the straightforward case of discrete time.

3.6 Theorem. *Let F be a bounded predictable process and let $X = \int F\, dM$. If M is a martingale, then so is X. If M is a submartingale and F is positive, then X is a submartingale.*

Proof. Suppose that M is a martingale. Since $F_0, \dots, F_n$ and $M_0, \dots, M_n$ are in $\mathcal{F}_n$, so is X_n; that is, X is adapted to $\mathcal{F}$. Since F is bounded, say by some constant $b > 0$, we see that $|X_n|$ is bounded by b times $|M_0| + |M_1 - M_0| + \cdots + |M_n - M_{n-1}|$, which is integrable. So X is integrable. Finally,

$$\begin{aligned} \mathbb{E}_n\ (X_{n+1} - X_n) &= \mathbb{E}_n\ (M_{n+1} - M_n) F_{n+1} \\ &= F_{n+1}\ \mathbb{E}_n\ (M_{n+1} - M_n) = F_{n+1} \cdot 0 = 0, \end{aligned}$$

where the second equality uses the predictability of F to move F_{n+1} outside the conditional expectation $\mathbb{E}_n$, and the third equality is merely the martingale equality for M. Hence, X is a martingale. If M is a submartingale and F is positive, the third and fourth equalities become $\geq$, and X is a submartingale. □

3.7 Heuristics. Here is an interpretation of the preceding theorem. A person buys and sells shares of a certain stock, presumably to make a profit. Let M_n be the price of a share at time n, and let F_n denote the number of shares owned during the time interval $(n-1, n]$. Then, the profit made during $(n-1, n]$ is $(M_n - M_{n-1}) \cdot F_n$. Hence, in 3.4, X_n is the sum of the initial value $X_0 = M_0 F_0$ and the total profit made during $(0, n]$. Since the decision on how many shares to own during $(n, n+1]$ must be made at time n based on the information $\mathcal{F}_n$ available at that time, it follows that F_{n+1} be $\mathcal{F}_n$-measurable, that is, F be predictable. For reasons mentioned in Example 2.6, the price process M should be a martingale. Then, the preceding theorem shows that it is impossible to make a profit systematically (or to lose systematically); no matter what "strategy" F one uses, X has no systematic tendency to move up or down.

Predictability

To enhance the value of the preceding theorem, the following provides some examples of predictable processes. Other examples may be constructed by noting that the class of predictable processes form a linear space that is closed under all limits.

3.8 Example. Let S and T be stopping times of $\mathcal{F}$ with $S \leq T$. Let V be a random variable in $\mathcal{F}_S$. Then,

$$V\ 1_{(S,T]}, \quad V\ 1_{(S,\infty]}, \quad 1_{(S,T]}, \quad 1_{[0,T]}$$

are all predictable processes. To see these, we start with the second, with $F = V\ 1_{(S,\infty]}$, that is, $F_n = V \cdot 1_{(S,\infty]}(n)$. Note that,

$$F_{n+1} = V\ 1_{\{S<n+1\}} = V \cdot 1_{\{S\leq n\}}\ \in \mathcal{F}_n$$

by Theorem 1.10 (or Theorem 1.16d with $T = n$). Thus, $V\ 1_{(S,\infty]}$ is predictable. Since $V \in \mathcal{F}_S$, and $\mathcal{F}_S \subset \mathcal{F}_T$ by the hypothesis that $S \le T$, we have $V \in \mathcal{F}_T$, and the preceding sentence implies that $V\ 1_{(T,\infty]}$ is predictable. Hence, the difference of the two, $V 1_{(S,T]}$, is predictable. Taking $V = 1$ shows that $1_{(S,T]}$ is predictable. Taking $T = \infty$, we see that $1_{(S,\infty]}$ is predictable, and finally, $1_{[0,S]} = 1 - 1_{(S,\infty]}$ is predictable.

Martingales stopped

Let $M = (M_n)$ be a process. Let T be a random time with values in $\bar{\mathbb{N}}$. Then, the process X defined by

3.9 $$X_n(\omega) = M_{n \wedge T(\omega)}(\omega) = \begin{cases} M_n(\omega) & \text{if } n \le T(\omega) \\ M_{T(\omega)}(\omega) & \text{if } n > T(\omega) \end{cases}$$

is called the process M stopped at T. We observe that X is the integral 3.5 with $F = 1_{[0,T]}$. This F is bounded and positive, and further, it is predictable when T is a stopping time (see the preceding example). Hence, the following is immediate from Theorem 3.6.

3.10 Theorem. *Let T be a stopping time. Let X be the process M stopped at T. If M is a martingale, then so is X. If M is a submartingale, then so is X.*

Doob's stopping theorem

This theorem captures the essence of the martingale property. For a martingale, given the cumulative information available at present, our estimate of any future increment is zero. Doob's theorem enables us to take the present and future times to be stopping times with some restriction (see 4.12, 4.13, and 5.8 as well), and further, it adds a simpler, more intuitive, characterization of the martingale property. The time set is still $\mathbb{N}$.

3.11 Theorem. *Let M be adapted to $\mathcal{F}$. Then, the following are equivalent:*

a) M is a submartingale.

b) For every pair of bounded stopping times S and T with $S \le T$, the random variables M_S and M_T are integrable and

3.12 $$\mathbb{E}_S(M_T - M_S) \ge 0.$$

c) For every pair of bounded stopping times S and T with $S \le T$, the random variables M_S and M_T are integrable and

3.13 $$\mathbb{E}\,(M_T - M_S) \ge 0.$$

These statements remain equivalent when (a) is changed to read "M is a martingale" provided that the inequalities in 3.12 and 3.13 are changed to equalities.

3.14 Remark. In fact, when (a) is changed to read "M is a martingale", then the inequality 3.12 needs to be replaced by equality, and (c) can be replaced with the following:

c) For every bounded stopping time T, the random variable M_T is integrable and $\mathbb{E}\ M_T = \mathbb{E}\ M_0$.

Proof. The martingale case is immediate from the submartingale case, since M is a martingale if and only if both M and $-M$ are submartingales. We shall, therefore, show that (a)⇒(b)⇒(c)⇒(a).

i) Let M be a submartingale. Let S and T be stopping times with $S(\omega) \le T(\omega) \le n$ for all ω, where n is some fixed integer. Let V be a bounded positive variable in $\mathcal{F}_S$. Putting $F = V\ 1_{(S,T]}$ in 3.4 yields a process X such that

$$X_n - X_0 = V \cdot (M_T - M_S).$$

The process F is predictable as noted in Example 3.8, and it is bounded and positive since V is so. Thus, by Theorem 3.6, the process X is a submartingale. The particular case with $V = 1$ and $S = 0$ shows that M_T is integrable (since X_n and X_0 are so), and the case with $V = 1$ and $T = n$ shows that M_S is integrable. Finally, recalling that $V \in \mathcal{F}_S$ and using the defining property for $\mathbb{E}_S$, we get

$$\mathbb{E}\ V\ \mathbb{E}_S(M_T - M_S) = \mathbb{E}\ V \cdot (M_T - M_S) = \mathbb{E}(X_n - X_0) \ge 0,$$

where the last inequality follows from the submartingale inequality for X. Since this holds for arbitrary V positive and bounded in $\mathcal{F}_S$, the random variable $\mathbb{E}_S(M_T - M_S)$ must be positive. This proves the implication (a)⇒(b).

ii) Suppose that (b) holds. Taking expectations on both sides of 3.12 yields 3.13, since $\mathbb{E}\ \mathbb{E}_S = \mathbb{E}$. So, (b)⇒(c).

iii) Suppose that (c) holds. Then, the integrability of M_n follows from that of M_T for the particular choice of $T = n$; and adaptedness of M is by hypothesis. Thus, to show (a), there remains to check the submartingale inequality $\mathbb{E}_m(M_n - M_m) \ge 0$, which is equivalent to showing that

3.15 $$\mathbb{E}\ 1_H\ \mathbb{E}_m(M_n - M_m) \ge 0, \qquad 0 \le m < n, H \in \mathcal{F}_m.$$

Fix m, n, H such. Define, for ω in Ω,

$$S(\omega) = m, \qquad T(\omega) = n 1_H(\omega) + m 1_{\Omega \setminus H}(\omega).$$

Now, S is a fixed time and is a stopping time trivially. Since $T \ge S$ and $H \in \mathcal{F}_S$, the time T is foretold at the time $S = m$; hence, T is a stopping time. Obviously, $S \le T \le n$. Finally, $M_T - M_S = 1_H \cdot (M_n - M_m)$ by the way T and S are defined. Now 3.13 shows that 3.15 holds as needed. □

Upcrossings

Let M be an adapted process. Fix a and b in $\mathbb{R}$ with $a < b$. Put $T_0 = -1$ for convenience and for each integer $k \geq 1$ define

3.16 $S_k = \inf \{ n > T_{k-1} : M_n \leq a \}, \quad T_k = \inf \{ n > S_k : M_n \geq b \}.$

Since M is adapted, $(S_1, T_1, S_2, T_2, \ldots)$ is an increasing sequence of stopping times; $S_1, S_2, \ldots$ are called the downcrossing times of the interval (a, b), and $T_1, T_2, \ldots$ are called the upcrossing times; See Figure 4 for an illustration. Then,

3.17 $$U_n(a, b) = \sum_{k=1}^{\infty} 1_{(0,n]} \circ T_k$$

is the number of upcrossings of (a, b) completed by M during $[0, n]$.

As in Heuristics 3.8, think of M_n as the price at time n of a share of some stock. Imagine someone who buys a share when the price hits a or below and sells it later when the price becomes b or above, repeating the scheme forever. Then, he buys a share at time S_1 and sells it at T_1, buys a share at time S_2 and sells it at T_2, and so on. The number of buy-sell cycles completed during $[0, n]$ is $U_n(a, b)$. The strategy employed is that of holding F_n shares during $(n - 1, n]$, where

3.18 $$F_n = \sum_{k=1}^{\infty} 1_{(S_k, T_k]}(n), \quad n \geq 1,$$

and we put $F_0 = 0$ for definiteness. Now $X = \int F dM$ describes the evolution of his capital, and $X_n - X_0$ is the profit during $(0, n]$, which profit is at least $(b - a) U_n(a, b)$ assuming that the share being held at time n (if any) is worth more than what it was bought for. This heuristic observation will be of use in the proof next.

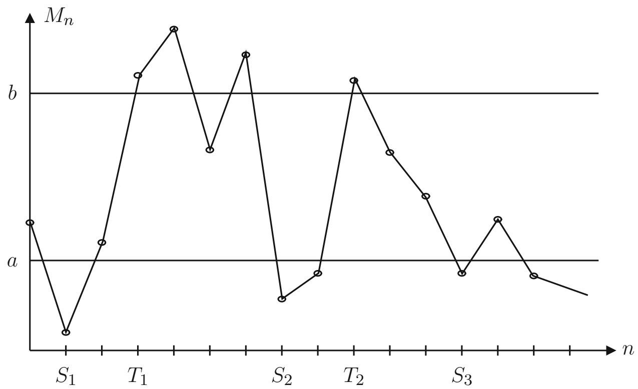


Figure 4: Upcrossing times of (a, b) are $T_1, T_2, \ldots$.

3.19 Proposition. *Suppose that M is a submartingale. Then,*

$$(b-a)\mathbb{E}\,U_n(a,b) \le \mathbb{E}\left[(M_n-a)^+ - (M_0-a)^+\right].$$

Proof. An upcrossing of (a,b) by M is the same as an upcrossing of $(0,b-a)$ by the process $(M-a)^+$, and the latter is again a submartingale by Remark 2.4g. Thus, we may and do assume that $a=0$ and $M \ge 0$.

Let $X = \int F dM$, defined by 3.4, with F given by 3.18. Note that F is predictable. Thus, $F_{k+1} \in \mathcal{F}_k$, and we have

$$\begin{aligned}\mathbb{E}_k(X_{k+1}-X_k) &= \mathbb{E}_k(M_{k+1}-M_k)F_{k+1} \\ &= F_{k+1}\mathbb{E}_k(M_{k+1}-M_k) \le \mathbb{E}_k(M_{k+1}-M_k),\end{aligned}$$

where the inequality follows from the positivity of $\mathbb{E}_k(M_{k+1}-M_k)$ and the observation that $F_{k+1} \le 1$. Taking expectations on both sides and summing over k we get

$$\mathbb{E}(X_n - X_0) \le \mathbb{E}(M_n - M_0).$$

On the other hand, as mentioned as a heuristic remark, $X_n - X_0 \ge bU_n(0,b)$ since $M_n \ge 0$ and $a=0$. Hence,

$$b\,\mathbb{E}\,U_n(0,b) \le \mathbb{E}(X_n-X_0) \le \mathbb{E}(M_n-M_0),$$

which is the claim when $a=0$ and $M \ge 0$. □

Maxima and minima

Let $M = (M_n)$ be a process adapted to $\mathcal{F}$. For n in $\mathbb{N}$, define

3.20 $$M_n^* = \max_{k\le n} M_k, \quad m_n^* = \min_{k \le n} M_k,$$

the maxima and minima.

3.21 Theorem. *Suppose that M is a submartingale. Then, for $b>0$,*

$$\begin{aligned}&b\,\mathbb{P}\{M_n^* \ge b\} \le \mathbb{E}\,M_n 1_{\{M_n^* \ge b\}} \le \mathbb{E}\,M_n^+, \\ &b\,\mathbb{P}\{m_n^* \le -b\} \le -\mathbb{E}\,M_0 + \mathbb{E}\,M_n 1_{\{m_n^* > -b\}} \le \mathbb{E}\,M_n^+ - \mathbb{E}\,M_0.\end{aligned}$$

3.22 Remark. It is convenient to think of these inequalities in terms of the stopping times (we suppress their dependence on b)

$$T = \inf\{n \ge 0 : M_n \ge b\}, \quad S = \inf\{n \ge 0 : M_n \le -b\},$$

that is, the time of first entrance to $[b,\infty)$ by M and the time of first entrance to $(-\infty,-b]$. Note that

3.23 $$\{M_n^* \ge b\} = \{T \le n\}, \quad \{m_n^* \le -b\} = \{S \le n\}.$$

We shall give the proof below in terms of T and S.

Proof. Fix b, fix n. Note that, on the set $\{T \le n\}$, we have $M_{T\wedge n} = M_T \ge b$. Thus,

$$b\ 1_{\{T\le n\}} \le M_{T\wedge n} 1_{\{T\le n\}} \le (\mathbb{E}_{T\wedge n} M_n) 1_{\{T\le n\}} = \mathbb{E}_{T\wedge n} M_n 1_{\{T\le n\}},$$

where the second inequality is Doob's submartingale inequality 3.12 applied with the bounded stopping times $T \wedge n$ and n, and the last equality uses Theorem 1.16d to the effect that $\{T \le n\} \in \mathcal{F}_{T\wedge n}$. Taking expectations on both sides yields the first inequality concerning M_n^* in view of Remark 3.22; the second inequality is obvious.

Similarly, on the set $\{S \le n\}$, we have $M_S \le -b$ and, hence,

$$M_{S\wedge n} = M_S\ 1_{\{S\le n\}} + M_n\ 1_{\{S>n\}} \le -b\ 1_{\{S\le n\}} + M_n\ 1_{\{S>n\}}.$$

Taking expectations, and noting that $\mathbb{E}\ M_0 \le \mathbb{E}\ M_{S\wedge n}$ by the submartingale inequality 3.13 applied with the bounded stopping times 0 and $S \wedge n$, we obtain the first inequality claimed for m_n^*. The second is obvious. □

When M is a martingale, $|M|^p$ is a submartingale for $p \ge 1$ provided that M_n be in L^p for every n. Then, applying the preceding theorem to the submartingale $|M|^p$ yields the following corollary. This is called Doob-Kolmogorov inequality; it is a generalization of Kolmogorov's inequality (Lemma III.7.1) for sums of independent variables.

3.24 COROLLARY. *Let M be a martingale in L^p for some p in $[1, \infty)$. Then, for $b > 0$,*

$$b^p\ \mathbb{P}\{\ \max_{k\le n} |M_k| > b\ \} \le \mathbb{E}\ |M_n|^p.$$

Another corollary, this time about the submartingale M directly, can be obtained by combining the two statements of Theorem 3.21:

3.25 $$b\,\mathbb{P}\{\ \max_{k\le n} |M_k| > b\ \} \le 2\,\mathbb{E}\,M_n^+ - \mathbb{E}\,M_0\ \le\ 3\ \max_{k\le n} \mathbb{E}\,|M_k|\,.$$

The following gives a bound on the expected value of the maxima of $|M|$ when M is a martingale. It is called *Doob's norm inequality.*

3.26 THEOREM. *Let M be a martingale in L^p for some $p > 1$. Let q be the exponent conjugate to p, that is, $1/p + 1/q = 1$. Then,*

$$\mathbb{E}\ \max_{k\le n} |M_k|^p \le q^p\ \mathbb{E}\ |M_n|^p.$$

Proof. Fix n, and introduce $Z = \max_{k\le n} |M_k|$ for typographical ease. We are to show that

3.27 $$\mathbb{E}\ Z^p \le q^p\ \mathbb{E}\ |M_n|^p.$$

We start by noting that

$$Z^p = \int_0^Z dx\ px^{p-1} = \int_0^\infty dx\ px^{p-2} x\ 1_{\{Z \geq x\}},$$

and

$$\mathbb{E}\ x\ 1_{\{Z\geq x\}} = x\ \mathbb{P}\{\max_{k\leq n} |M_k| \geq x\} \leq \mathbb{E}\ |M_n| \cdot 1_{\{Z\geq x\}}$$

by Theorem 3.21 applied to the submartingale $|M|$. Thus,

$$\begin{aligned} Ebb\ Z^p &\leq \mathbb{E}\ |M_n| \int_0^\infty dx\ px^{p-2}\ 1_{\{Z\geq x\}} = \mathbb{E}\ |M_n|\ q\ Z^{p-1} \\ &\leq q\ (\mathbb{E}\ |M_n|^p)^{1/p}\ (\mathbb{E}\ Z^p)^{1/q}, \end{aligned}$$

where the last inequality follows from Hölder's, II.3.6a. Solving this for $\mathbb{E}\ Z^p$ yields the desired bound 3.27. □

Exercises

3.28 *Doob's Decomposition.* Let $X = (X_n)$ be a submartingale and let

$$X = X_0 + M + A$$

be its Doob decomposition as in Theorem 3.2. Show that X is L^1-bounded if and only if both M and A are L^1-bounded.

3.29 *Martingales in L^2.* Let M be a martingale in L^2 and let Q by its quadratic variation process; see 2.30 and 2.33. Show that the martingale $X = M^2 - M_0^2 - Q$ has the form (see 3.5)

$$X = \int F\ dM$$

with $F_0 = 0$ and $F_n = 2M_{n-1}$, $n \geq 1$.

3.30 *Continuation.* Note that Q is a submartingale with $Q_0 = 0$. Let $Q = Y + A$ be its Doob decomposition with Y a martingale and A an increasing predictable process. Describe Y and A. Show that, with $N = X + Y$,

$$M^2 = M_0^2 + N + A$$

is Doob's decomposition for the submartingale M^2.

3.31 *Upcrossings.* Recall the definitions 3.16-3.18. Show that F can be obtained recursively by, starting with $F_0 = 0$,

$$F_{n+1} = F_n\ 1_{\{M_n<b\}} + (1 - F_n)\ 1_{\{M_n\leq a\}}, \quad n \geq 0.$$

Define the stopping times $S_1, T_1, S_2, T_2, \ldots$ in terms of the F_n. Show that

$$U_n(a, b) = \sum_{k=1}^n F_k \cdot (1 - F_{k+1}) = \sum_{k=1}^n F_k\ 1_{\{M_k\geq b\}}.$$

4 Martingale Convergence

This section is on the fundamental results of the classical theory of martingales. We give the basic convergence theorems, characterization of uniformly integrable martingales, and a sample of applications: Hunt's extension of the dominated convergence theorem, Lévy's extension of the Borel-Cantelli lemma, new proofs of Kolmogorov's 0-1 law and the strong law of large numbers, and a constructive proof of the Radon-Nikodym theorem.

As usual, $(\Omega, \mathcal{H}, \mathbb{P})$ is the probability space in the background. The index set is $\mathbb{N}$ unless stated otherwise. The filtration $\mathcal{F}$ is kept in the background as well, and Convention 1.21 on conditional expectations is in force throughout. All martingales, stopping times, and so on are relative to the filtration $\mathcal{F}$.

Martingale convergence theorem

The next theorem is basic. Let X be a submartingale. Doob's decomposition of it shows that it has a tendency to increase. If that tendency can be curbed, then it should be convergent.

4.1 Theorem. *Let X be a submartingale. Suppose that*

4.2 $$\sup_n \mathbb{E}\, X_n^+ < \infty.$$

Then, the sequence (X_n) converges almost surely to an integrable random variable.

4.3 Remarks. a) The condition 4.2 is that the process X^+ be L^1-bounded. Since X is a submartingale, so is X^+ by Remark 2.4g, and $\mathbb{E}\, X_n^+$ is increasing in n as a result. The condition 4.2 delimits the upward tendency of X, and the convergence of X becomes intuitive.

b) If X is a negative submartingale, then 4.2 is automatic and X converges.

c) If X is a positive supermartingale, then the preceding remark applies to $-X$ and, hence, X converges.

d) Since every martingale is a submartingale and a supermartingale, the preceding theorem and remarks apply: If X is a martingale, and if X is positive or negative or bounded from below by an integrable random variable or bounded from above similarly, then X converges almost surely to an integrable random variable X_∞.

e) When X is a submartingale, $\mathbb{E}\, X_n \geq \mathbb{E}\, X_0$ for every n, and

$$\mathbb{E}\, X_n^+ \leq \mathbb{E}\, |X_n| = 2\, \mathbb{E}\, X_n^+ - \mathbb{E}\, X_n \leq 2\, \mathbb{E}\, X_n^+ - \mathbb{E}\, X_0.$$

Hence, the condition that X^+ be L^1-bounded is equivalent to requiring that X be L^1-bounded, that is, 4.2 holds if and only if

4.4 $$\sup_n \mathbb{E}\, |X_n| < \infty.$$

Proof. Pick an outcome ω, suppose that the sequence of numbers $X_n(\omega)$ does not have a limit; then its limit inferior is strictly less than its limit superior, in which case there are at least two rationals a and b with $a < b$ that can be inserted between the two limits, and which in turn implies that the sequence upcrosses the interval (a, b) infinitely often. The set of all such ω is the union, over all rationals a and b with $a < b$, of the sets $\{U(a,b) = +\infty\}$, where $U(a,b) = \lim_n U_n(a,b)$, the total number of upcrossings of (a,b). Thus, to show that $\lim X_n$ exists almost surely, it is enough to show that for every pair of rationals a and b with $a < b$ we have $U(a,b) < \infty$ almost surely.

Fix $a < b$ such. Since $U_n(a,b)$ is increasing in n,

$$(b-a)\mathbb{E}\,U(a,b) = (b-a)\lim \mathbb{E}\,U_n(a,b) \le \sup \mathbb{E}(X_n - a)^+ \le \sup \mathbb{E}X_n^+ + |a| < \infty,$$

where we used, in succession, the monotone convergence theorem, Proposition 3.19 on upcrossings, the observation that $(x-a)^+ \le x^+ + |a|$, and the condition 4.2. Thus, $U(a,b) < \infty$ almost surely.

It follows that $X_\infty = \lim X_n$ exists almost surely. By Fatou's lemma and Remark 4.3e,

$$\mathbb{E}\,|X_\infty| = \mathbb{E}\,\liminf |X_n| \le \liminf \mathbb{E}|X_n| \le 2\sup_n \mathbb{E}\,X_n^+ - \mathbb{E}\,X_0 < \infty,$$

which shows that the limit is integrable (and thus real-valued), and hence, X is convergent. □

Convergence and uniform integrability

The following improves upon the preceding theorem in the presence of uniform integrability. Recall 1.8 *et seq.* on extending the filtration $\mathcal{F}$ onto $\bar{\mathbb{N}}$ by setting $\mathcal{F}_\infty = \lim \mathcal{F}_n = \vee_n \mathcal{F}_n$.

4.5 THEOREM. *Let X be a submartingale. Then, X converges almost surely and in L^1 if and only if it is uniformly integrable. Moreover, if it is so, setting $X_\infty = \lim X_n$ extends X to a submartingale $\bar{X} = (X_n)_{n\in\bar{\mathbb{N}}}$.*

Proof. If X converges almost surely and in L^1, then it must be uniformly integrable; see Theorem III.4.6.

If the submartingale X is uniformly integrable, then it is L^1-bounded by Remark II.3.13c and the condition 4.2 follows from Remark 4.3e. Thus, X converges almost surely by Theorem 4.1, and also in L^1 by Theorem III.4.6. Moreover, then, the limit X_∞ is integrable by 4.1 and belongs to $\mathcal{F}_\infty$ since all the X_n belong to $\mathcal{F}_\infty$. To show that $\bar{X}$ is a submartingale over $\bar{\mathbb{N}}$, there remains to show that, for every m in $\mathbb{N}$,

4.6 $$\mathbb{E}_m(X_\infty - X_m) \ge 0.$$

Fix m. Fix H in $\mathcal{F}_m$. The submartingale inequality for X implies that

$$\mathbb{E}\,1_H \cdot (X_n - X_m) = \mathbb{E}\,1_H\,\mathbb{E}_m(X_n - X_m) \ge 0$$

for every $n \geq m$. Thus, since $X_n - X_m$ goes to $X_\infty - X_m$ in L^1 as $n \to \infty$,

$$\mathbb{E}\ 1_H(X_\infty - X_m) = \lim_n\ \mathbb{E}\ 1_H(X_n - X_m) \geq 0$$

by Proposition III.4.7. Since H in $\mathcal{F}_m$ is arbitrary, this implies 4.6. □

Uniformly integrable martingales

The following theorem characterizes uniformly integrable martingales and identifies their limits. Its proof is nearly immediate from Proposition 2.7 and the preceding theorem.

4.7 THEOREM. *A process $M = (M_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale if and only if*

4.8 $$M_n = \mathbb{E}_n\ Z, \qquad n \in \mathbb{N},$$

for some integrable random variable Z. If so, it converges almost surely and in L^1 to the integrable random variable

4.9 $$M_\infty = \mathbb{E}_\infty\ Z,$$

and, moreover, $\bar{M} = (M_n)_{n\in\bar{\mathbb{N}}}$ is again a uniformly integrable martingale.

Proof. If M has the form 4.8, then it is a uniformly integrable martingale as was shown in Proposition 2.7. If M is a uniformly integrable martingale, then the preceding theorem shows that it converges almost surely and in L^1 to some integrable random variable M_∞ and that $\bar{M} = (M_n)_{n\in\bar{\mathbb{N}}}$ is again a martingale; it follows from the martingale property for $\bar{M}$ that M has the form 4.8 with $Z = M_\infty$. This completes the proof of the first statement and much of the second.

To complete the proof, there remains to show that if 4.8 holds then 4.9 holds as well, which amounts to showing that

4.10 $$\mathbb{E}\ M_\infty\ 1_H = \mathbb{E}\ Z\ 1_H$$

for every H in $\mathcal{F}_\infty$. Let $\mathcal{D}$ be the collection of all H in $\mathcal{F}_\infty$ for which 4.10 holds. Then $\mathcal{D} \supset \mathcal{F}_n$ for each n since $\mathbb{E}_n\ M_\infty = M_n = \mathbb{E}_n\ Z$; thus, $\mathcal{D} \supset \cup_n \mathcal{F}_n$. Since $\mathcal{D}$ is clearly a d-system, and since it contains the p-system $\cup_n \mathcal{F}_n$, it follows that $\mathcal{D} \supset \mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$; this is by the monotone class theorem. So, 4.10 holds for every H in $\mathcal{F}_\infty$. □

4.11 COROLLARY. *For every integrable variable Z,*

$$\mathbb{E}_n\ Z \longrightarrow \mathbb{E}_\infty\ Z$$

almost surely and in L^1.

Proof is not needed; this is a partial re-statement of the preceding theorem. Note that, in particular, if $Z \in \mathcal{F}_\infty$ then $\mathbb{E}_n\, Z \to Z$; that is, if Z is revealed by the end of time, then our estimate of it at time n converges to it as $n \to \infty$.

The following supplements the preceding results and removes the boundedness condition in Doob's stopping theorem 3.11. In view of Theorem 4.7 above, the condition of the next theorem is equivalent to saying that $(M_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale and $M_\infty = \lim M_n$.

4.12 THEOREM. *Suppose that, for some integrable random variable Z,*

$$M_n = \mathbb{E}_n\, Z, \quad n \in \bar{\mathbb{N}}.$$

Then, for every stopping time T,

$$M_T = \mathbb{E}_T\, Z.$$

Moreover, for arbitrary stopping times S and T,

$$\mathbb{E}_S\, M_T = M_{S\wedge T}.$$

4.13 REMARKS. a) *On the meaning of M_T*: Since $M_n(\omega)$ is well-defined for every integer n and $n = +\infty$, the random variable M_T is well-defined even for ω with $T(\omega) = +\infty$.

b) *Doob's stopping theorem.* According to the first claim, M_T is the conditional expectation of Z given $\mathcal{F}_T$. Since Z is integrable, this implies that M_T is integrable. So, if S and T are arbitrary stopping times (taking values in $\bar{\mathbb{N}}$) with $S \le T$, the random variables M_S and M_T are integrable and

$$\mathbb{E}_S\, M_T = M_S$$

by the second claim. Thus, for uniformly integrable martingales, Doob's stopping theorem 3.11 remains true without the condition of boundedness on the stopping times.

Proof. We shall be using the repeated conditioning property, $\mathbb{E}_S\mathbb{E}_T = \mathbb{E}_{S\wedge T}$, a number of times without further comment. To prove the first claim, we start by noting that, for each n in $\mathbb{N}$,

$$M_{T\wedge n} = \mathbb{E}_{T\wedge n}\, M_n;$$

this follows from Doob's stopping theorem 3.11 for the martingale M used with the bounded stopping times $T \wedge n$ and n. Replacing M_n by $\mathbb{E}_n Z$, noting that $\mathbb{E}_{T\wedge n}\mathbb{E}_n = \mathbb{E}_{T\wedge n} = \mathbb{E}_n\mathbb{E}_T$, we get

$$M_{T\wedge n} = \mathbb{E}_n\mathbb{E}_T Z, \quad n \in \mathbb{N}.$$

As $n \to \infty$, the left side converges to M_T almost surely, whereas the right side converges to $\mathbb{E}_\infty\mathbb{E}_T Z = \mathbb{E}_T Z$ by Corollary 4.11 applied to the integrable

random variable $\mathbb{E}_T Z$. Thus, $M_T = \mathbb{E}_T Z$ as claimed. The second claim is immediate:

$$\mathbb{E}_S M_T = \mathbb{E}_S \mathbb{E}_T Z = \mathbb{E}_{S\wedge T} Z = M_{S\wedge T}. \qquad \square$$

The following corollary is immediate from the preceding theorem and Lemma 2.8.

4.14 Corollary. *If $(M_n)_{n\in\bar{\mathbb{N}}}$ is a uniformly integrable martingale, then the collection*

$$\{M_T : \ T \text{ is a stopping time}\}$$

is uniformly integrable. $\square$

Within the proof of the last theorem, we have shown that

4.15 $$\mathbb{E}_{T\wedge n}\, Z \longrightarrow \mathbb{E}_T\, Z$$

almost surely and in L^1; this follows from applying Corollary 4.11 to $\mathbb{E}_n \mathbb{E}_T Z = \mathbb{E}_{T\wedge n} Z$. Here is a consequence of this useful fact.

4.16 Proposition. *Suppose that $(\Omega, \mathcal{H}, \mathbb{P})$ is complete, and all negligible events belong to $\mathcal{F}_0$ (and therefore to all the $\mathcal{F}_n$). Then, for every stopping time T of the filtration $(\mathcal{F}_n)$,*

$$\mathcal{F}_T = \lim_n \mathcal{F}_{T\wedge n} = \vee_n \mathcal{F}_{T\wedge n}.$$

Proof. Let $\hat{\mathcal{F}}_n = \mathcal{F}_{T\wedge n}$; we are to show that $\hat{\mathcal{F}}_\infty = \mathcal{F}_T$. Since $\hat{\mathcal{F}}_n \subset \mathcal{F}_T$ for every n, we have $\hat{\mathcal{F}}_\infty \subset \mathcal{F}_T$. To show the converse containment, let Z be a bounded variable in $\mathcal{F}_T$. Then, $Z = \mathbb{E}_T Z$ by definition and, thus, is the almost sure limit of $\mathbb{E}_{T\wedge n} Z \in \hat{\mathcal{F}}_n \subset \hat{\mathcal{F}}_\infty$. Since $\hat{\mathcal{F}}_\infty \supset \mathcal{F}_0$ and $\mathcal{F}_0$ includes every negligible event, it follows that $Z \in \hat{\mathcal{F}}_\infty$. So, $\mathcal{F}_T \subset \hat{\mathcal{F}}_\infty$ as well. $\square$

Convergence in reversed time

In this subsection, the index set is $\mathbb{T} = \{\ldots, -2, -1, 0\}$, and $\mathcal{F}$ is a filtration on $\mathbb{T}$, that is, $\mathcal{F}_m \subset \mathcal{F}_n$ for $m < n$ as before but for m and n in $\mathbb{T}$.

4.17 Theorem. *Let $X = (X_n)_{n\in\mathbb{T}}$ be a martingale relative to $\mathcal{F}$. Then, X is uniformly integrable and, as $n \to -\infty$, it converges almost surely and in L^1 to the integrable random variable $X_{-\infty} = \mathbb{E}_{-\infty} X_0$, where $\mathcal{F}_{-\infty} = \cap_{n\in\mathbb{T}} \mathcal{F}_n$.*

Proof. i) The martingale property for X implies that

$$X_n = \mathbb{E}_n\, X_0, \qquad n \in \mathbb{T}.$$

Thus, X has the same form as in Proposition 2.7 and is uniformly integrable as shown there.

ii) Let a and b be real numbers with $a < b$. By Proposition 3.19 on upcrossings applied to the martingale $(X_n, X_{n+1}, \ldots, X_0, X_0, X_0, \ldots)$, the expected number of upcrossings of (a, b) by X during $[n, 0]$ is bounded by

$$\frac{1}{b-a}\mathbb{E}\ [(X_0 - a)^+ - (X_n - a)^+] \le \frac{1}{b-a}\ \mathbb{E}\ (X_0 - a)^+ < \infty.$$

Thus, the number of upcrossings of (a, b) by X over $(-\infty, 0]$ is almost surely finite, just as in the proof of the martingale convergence theorem 4.1. Hence, as in 4.1 again, X converges almost surely to some integrable random variable $X_{-\infty}$ as $n \to -\infty$. By the uniform integrability of X, the convergence is in L^1 as well. Clearly, the limit belongs to $\mathcal{F}_n$ for every n in $\mathbb{T}$, and hence, is in $\mathcal{F}_{-\infty}$. □

The next corollary mirrors Corollary 4.11. No proof is needed.

4.18 Corollary. *For every integrable random variable Z,*

$$\mathbb{E}_n\ Z \longrightarrow \mathbb{E}_{-\infty}\ Z$$

almost surely and in L^1 as $n \to -\infty$.

The preceding proof extends to submartingales on $\mathbb{T}$, but requires a condition to check the downward tendency of the submartingale as n goes toward $-\infty$.

4.19 Theorem. *Let $X = (X_n)_{n\in\mathbb{T}}$ be a submartingale relative to $\mathcal{F}$. Suppose that*

4.20 $$\inf_n\ \mathbb{E}\ X_n > -\infty.$$

Then, X is uniformly integrable and, as $n \to -\infty$, converges almost surely and in L^1 to an integrable random variable $X_{-\infty}$.

Proof. i) Since X is a submartingale, $\mathbb{E}\ X_n$ decreases as n decreases, and 4.20 implies that $a = \lim\ \mathbb{E}\ X_n$ is finite. Fix $\varepsilon > 0$, take the negative integer m such that $\mathbb{E}\ X_n \le a + \varepsilon$ for all $n \le m$, and recall that X^+ is again a submartingale. Thus, for $n \le m$ and $H \in \mathcal{F}_n$,

$$-\mathbb{E}\ X_n \le -\mathbb{E}\ X_m + \varepsilon, \quad \mathbb{E}\ X_n^+\ 1_H \le \mathbb{E}\ X_m^+\ 1_H, \quad \mathbb{E}\ X_n\ 1_{\Omega\setminus H} \le \mathbb{E}\ X_m\ 1_{\Omega\setminus H}.$$

These inequalities imply, since $|Z|\ 1_H = -Z + 2Z^+ 1_H + Z\ 1_{\Omega\setminus H}$ for all variables Z, that

4.21 $$\mathbb{E}\ |X_n|\ 1_H \le \mathbb{E}\ |X_m|\ 1_H + \varepsilon$$

for all $n \le m$ and H in $\mathcal{F}_n$.

Next, fix $b > 0$, take $H = \{\ |X_n| > b\}$, and use Markov's inequality and the submartingale property for X^+ and the fact that $\mathbb{E}\ X_n \ge a$ for all n. We get

$$b\ \mathbb{P}(H) \le \mathbb{E}\ |X_n| \le 2\mathbb{E}\ X_n^+ - \mathbb{E}\ X_n \le 2\mathbb{E}\ X_0^+ - a,$$

which shows that the probability here can be made as small as desired by taking b large enough. Using this in 4.21 and recalling the integrability of X_m, we deduce that $(X_n)_{n \le m}$ is uniformly integrable. Since adding the integrable random variables $X_{m+1}, X_{m+2}, \ldots, X_0$ does not affect the uniform integrability, we conclude that X is uniformly integrable.

ii) The remainder of the proof follows the part (ii) of the proof of 4.17 word for word. □

In the remainder of this section we give a sample of applications and extensions of the convergence theorems above.

Hunt's dominated convergence theorem

This is a useful extension of the dominated convergence theorem. Note that there is no assumption of adaptedness for the sequence.

4.22 Theorem. *Let (X_n) be dominated by an integrable random variable and suppose that $X_\infty = \lim X_n$ exists almost surely. Then, the sequence $(\mathbb{E}_n\ X_n)$ converges to $\mathbb{E}_\infty\ X_\infty$ almost surely and in L^1.*

Proof. Suppose that $|X_n| \le Z$ for every n, where Z is integrable. Then, (X_n) is uniformly integrable, its limit X_∞ is integrable, and Corollary 4.11 implies that $\mathbb{E}_n X_\infty \to \mathbb{E}_\infty X_\infty$ almost surely and in L^1. Thus, the proof is reduced to showing that, as $n \to \infty$,

$$|\mathbb{E}_n X_n - \mathbb{E}_n X_\infty| \to 0$$

almost surely and in L^1. Convergence in L^1 is easy:

$$\mathbb{E}\ |\mathbb{E}_n X_n - \mathbb{E}_n X_\infty| \le \mathbb{E}\ \mathbb{E}_n\ |X_n - X_\infty| = \mathbb{E}\ |X_n - X_\infty| \to 0$$

since $X_n \to X_\infty$ in L^1 as well. To show the almost sure convergence, let $Z_m = \sup_{n \ge m} |X_n - X_\infty|$. Observe that, almost surely,

$$\limsup_{n\to\infty} |\mathbb{E}_n X_n - \mathbb{E}_n X_\infty| \le \limsup_{n\to\infty} \mathbb{E}_n Z_m = \mathbb{E}_\infty Z_m,$$

where the last equality follows from Corollary 4.11 after noting that Z_m is integrable, in fact, $|Z_m| \le 2Z$. This completes the proof since $\mathbb{E}_\infty Z_m \to \mathbb{E}_\infty \lim Z_m = 0$ as $m \to \infty$ by the dominated convergence property IV.1.8 for conditional expectations. □

An extension of the Borel-Cantelli lemma

Let $X = (X_n)$ be a sequence, adapted to $(\mathcal{F}_n)$, of positive integrable random variables. Put $S_0 = 0$ and

$$S_n = S_0 + X_1 + \cdots + X_n, \quad n \in \mathbb{N}.$$

Then, $S = (S_n)_{n\in\mathbb{N}}$ is an increasing integrable process adapted to $\mathcal{F}$, that is, an increasing submartingale. Let

4.23 $$S = M + A$$

be its Doob decomposition: M is a martingale and A is an increasing predictable process. Since both S and A are increasing, the limits $S_\infty = \lim S_n$ and $A_\infty = \lim A_n$ are well-defined.

4.24 Proposition. *For almost every ω,*

$$A_\infty(\omega) < \infty \Rightarrow S_\infty(\omega) < \infty.$$

If X is bounded by some constant, then the reverse implication holds as well.

4.25 Remark. Recall that, in Doob's decomposition 4.23, we have

$$A_n = \mathbb{E}_0 X_1 + \mathbb{E}_1 X_2 + \cdots + \mathbb{E}_{n-1} X_n.$$

If the X_n are independent, then $A_n = \mathbb{E}S_n$, and the preceding proposition relates the convergence of S_n to the convergence of its mean, which relation is what the Borel-Cantelli lemma is about.

Proof. i) For b in $(0,\infty)$ let

$$T = \inf\{n : A_{n+1} > b\},$$

and let N be the martingale M stopped at T (see 3.9). Predictability of A implies that T is a stopping time, and Theorem 3.10 shows that N is a martingale. Since $M = S - A \geq -A$ and since $A_n \leq b$ on $\{n \leq T\}$, the martingale $N + b$ is positive, and, hence, it is almost surely convergent by Remark 4.3d. Hence, N is convergent almost surely.

ii) Let Ω_b be the almost sure event on which the limit N_∞ exists and is finite, and let $H_b = \Omega_b \cap \{A_\infty \leq b\}$. For every ω in H_b, we have $T(\omega) = \infty$, which means that $M_n(\omega) = N_n(\omega)$ for every integer n, which implies that the limit $M_\infty(\omega)$ of $M_n(\omega)$ exists and is finite, which in turn allows us to conclude that $S_\infty(\omega) = M_\infty(\omega) + A_\infty(\omega) < \infty$. Hence, on the event $H = H_1 \cup H_2 \cup \cdots$, we have $A_\infty < \infty$ and $S_\infty < \infty$, and noting that $\{A_\infty < \infty\} \setminus H$ is negligible completes the proof of the first statement.

iii) Next, suppose that the sequence X is bounded by some constant c. For b in $(0,\infty)$ define

$$T = \inf\{n : S_n > b\},$$

and let N be the martingale M stopped at T. Since $X_n \leq c$ for every n, the process N^+ is bounded by $b + c$. So, N converges almost surely to some finite random variable N_∞; this is by the martingale convergence theorem 4.1. The remainder of the proof of the second statement follows the part (ii) above with the letter A replaced by S, and S by A. □

Kolmogorov's 0-1 law

This is to illustrate the power of the martingale machinery by giving a short proof of Kolmogorov's 0-1 law. Let $X_1, X_2, \ldots$ be independent random variables. Suppose that $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Put $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$, and let $\mathcal{T} = \cap_n \mathcal{T}_n$, the tail σ-algebra.

4.26 PROPOSITION. *If $H \in \mathcal{T}$ then $\mathbb{P}(H)$ is either 0 or 1.*

Proof. By Corollary 4.11, for every event H,

$$\mathbb{E}_n\, 1_H \to \mathbb{E}_\infty\, 1_H$$

almost surely. When $H \in \mathcal{T}$, since $\mathcal{T}$ is independent of $\mathcal{F}_n$, we have $\mathbb{E}_n\, 1_H = \mathbb{E}\, 1_H = \mathbb{P}(H)$. On the other hand, since $\mathcal{T}_n \subset \mathcal{F}_\infty$ for every n, we have $\mathcal{T} \subset \mathcal{F}_\infty$, which implies that $\mathbb{E}_\infty\, 1_H = 1_H$. Thus, $1_H(\omega)$ is equal to the number $\mathbb{P}(H)$ for almost every ω, which makes the latter either 0 or 1. □

Of course, consequently, for every random variable X in $\mathcal{T}$, there is constant c in $[-\infty, +\infty]$ such that $X(\omega) = c$ for almost every ω.

Strong law of large numbers

Let $X_1, X_2, \ldots$ be independent and identically distributed real-valued random variables with finite mean a. Then, by the strong law of large numbers,

4.27 $$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

converges almost surely to the constant a. Here is a martingale proof of this, which shows that the convergence is in L^1 as well.

Let $\mathcal{F}_{-n} = \sigma(\bar{X}_n, \bar{X}_{n+1}, \ldots)$, which is the same as the σ-algebra generated by $\bar{X}_n$ and the tail $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. Since the vector $(X_1, \ldots, X_n)$ is independent of $\mathcal{T}_n$, the conditional expectation $\mathbb{E}_{-n}X_k$ of X_k given $\mathcal{F}_{-n}$ depends only on $\bar{X}_n$. Since the distribution of that vector remains invariant under permutations of its entries, $\mathbb{E}_{-n}X_1 = \cdots = \mathbb{E}_{-n}X_n$. But the sum of these n things is equal to $\mathbb{E}_{-n}(X_1 + \cdots + X_n) = n\bar{X}_n$. Hence,

4.28 $$\bar{X}_n = \mathbb{E}_{-n}X_1, \quad n = 1, 2, \ldots.$$

Corollary 4.18 on reversed martingales applies to the right side:

4.29 $$\bar{X}_\infty = \lim \bar{X}_n$$

exists almost surely and in L^1. Convergence in L^1 helps to see that

$$\mathbb{E}\, \bar{X}_\infty = \lim \mathbb{E}\, \bar{X}_n = a.$$

On the other hand, 4.27 shows that

$$\bar{X}_\infty = \lim_n \frac{1}{n}(X_{k+1} + \cdots + X_{k+n}),$$

which shows that $\bar{X}_\infty$ belongs to $\mathcal{T}_k$ for every k and, hence, to the tail σ-algebra $\cap_k \mathcal{T}_k$. By Kolmogorov's 0-1 law, $\bar{X}_\infty$ is a constant over an almost sure event, and obviously that constant is its mean a.

Radon-Nikodym theorem

Our aim here is to give a constructive, and intuitive, proof of the Radon-Nikodym theorem and a very useful extension of it due to Doob.

First, a definition: A σ-algebra $\mathcal{G}$ on Ω is said to be *separable* if it is generated by some sequence (H_n) of subsets of Ω. Then, letting $\mathcal{F}_n = \sigma(H_1, \ldots, H_n)$, we obtain a filtration $(\mathcal{F}_n)$ such that

4.30 $$\mathcal{F}_\infty = \lim \mathcal{F}_n = \vee_n \mathcal{F}_n = \mathcal{G}.$$

Indeed, each $\mathcal{F}_n$ has only a finite number of elements, and we can easily find a finite partition $\mathcal{D}_n$ of Ω such that $\mathcal{F}_n = \sigma(\mathcal{D}_n)$, that is, each set H in $\mathcal{F}_n$ is the union of some finite number of elements of $\mathcal{D}_n$. Obviously, $\mathcal{D}_n$ gets more and more refined as n increases. So, $\mathcal{G}$ is a separable σ-algebra if and only if it is generated by a sequence $(\mathcal{D}_n)$ of finite partitions of Ω.

For example, if $\Omega = [0,1]$, its Borel σ-algebra $\mathcal{G}$ is separable: for $\mathcal{D}_n$ take the partition whose elements are $[0,a]$, $(a,2a]$, $(2a,3a]$, $\ldots$, $(1-a,1]$ with $a = 1/2^n$. This example is worth keeping in mind.

4.31 THEOREM. *Let $\mathcal{G}$ be a separable sub-σ-algebra of $\mathcal{H}$. Let Q be a finite measure on $(\Omega, \mathcal{G})$. Suppose that Q is absolutely continuous with respect to P, the latter being the restriction of $\mathbb{P}$ to $\mathcal{G}$. Then, there exists a positive random variable Z in $\mathcal{G}$ such that*

4.32 $$Q(H) = \int_H \mathbb{P}(d\omega) Z(\omega), \quad H \in \mathcal{G}.$$

4.33 REMARKS. a) Of course, the conclusion is that

$$Z = \frac{dQ}{dP},$$

that is, Z is a version of the Radon-Nikodym derivative of Q with respect to P.

b) If $\mathcal{H}$ is separable, or if $\mathcal{H}$ differs from a separable σ-algebra by a collection of negligible events, then the theorem remains true with $\mathcal{G} = \mathcal{H}$. In fact, in most situations in probability theory, this remark is applicable to $\mathcal{H}$.

c) In fact, the separability condition can be dropped: The claim of the theorem is true for arbitrary sub-σ-algebra $\mathcal{G}$ of $\mathcal{H}$ (See the notes for references).

Proof. We start by constructing a sequence of random variables (this is the intuitive part) and give the proof through a series of lemmas.

4.34 Construction. For each n, let $\mathcal{F}_n$ be the σ-algebra generated by a finite partition $\mathcal{D}_n$ of Ω such that the sequence $(\mathcal{F}_n)$ is a filtration and 4.30 holds.

For each ω in Ω, there is a unique element H of $\mathcal{D}_n$ such that $\omega \in H$, and then we define $X_n(\omega)$ to be the ratio $Q(H)/P(H)$; in other words,

4.35 $$X_n(\omega) = \sum_{H \in \mathcal{D}_n} \frac{Q(H)}{P(H)}\, 1_H(\omega), \qquad n \in \mathbb{N},\ \omega \in \Omega,$$

with the convention that $0/0 = 0$. Obviously, each X_n is positive and is in $\mathcal{F}_n$ and takes finitely many values, and

4.36 $$Q(H) = \mathbb{E}\, 1_H X_n, \quad H \in \mathcal{F}_n.$$

4.37 Lemma. *The process (X_n) is a positive martingale with respect to the filtration $(\mathcal{F}_n)$; it converges almost surely to an integrable positive random variable Z in $\mathcal{G}$.*

Proof. The positivity and adaptedness are obvious. Taking $H = \Omega$ in 4.36 shows that $\mathbb{E}\, X_n = Q(\Omega) < \infty$. To see the martingale property, let $H \in \mathcal{F}_n$. Then, $H \in \mathcal{F}_{n+1}$ as well, and 4.36 shows that

4.38 $$\mathbb{E}\, 1_H X_n = Q(H) = \mathbb{E}\, 1_H X_{n+1}.$$

This is another way of saying that $\mathbb{E}_n\, X_{n+1} = X_n$. Thus, X is a positive martingale. The remaining claim is immediate from the convergence theorem 4.1 and Remark 4.3d. □

4.39 Lemma. *For every $\varepsilon > 0$ there is $\delta > 0$ such that, for every event H in $\mathcal{G}$,*

$$P(H) \le \delta \Rightarrow Q(H) \le \varepsilon.$$

Proof. This is by the assumed absolute continuity of Q with respect to P. We show it by contradiction. Suppose that for some $\varepsilon > 0$ there is no such δ. Then, there must exist H_n in $\mathcal{G}$ such that

$$P(H_n) \le 1/2^n, \quad Q(H_n) > \varepsilon.$$

Define $H = \limsup H_n$, that is, 1_H is the limit superior of the indicators of the H_n. By Borel-Cantelli for the probability measure P we have $P(H) = 0$, whereas

$$Q(H) \ge \limsup Q(H_n) \ge \varepsilon$$

by Fatou's lemma applied with the finite measure Q. This contradicts the absolute continuity $(P(H) = 0 \Rightarrow Q(H) = 0)$. □

The following lemma completes the proof of Theorem 4.31.

4.40 Lemma. *The martingale (X_n) is uniformly integrable, and its limit Z is in $\mathcal{G}$ and satisfies 4.32.*

Proof. Pick $\varepsilon > 0$ and choose $\delta > 0$ as in the preceding lemma. Let $b = Q(\Omega)/\delta$ and $H = \{X_n > b\}$. Since

$$P(H) \leq \frac{1}{b}\ \mathbb{E}\ X_n = \frac{1}{b}\ Q(\Omega) = \delta,$$

we have (recall 4.36)

$$\mathbb{E}\ X_n\ 1_{\{X_n > b\}} = \mathbb{E}\ 1_H\ X_n = Q(H) \leq \varepsilon.$$

Thus, (X_n) is uniformly integrable. By Lemma 4.37 it is a positive martingale and converges to Z in $\mathcal{G}$ almost surely. Hence, it converges to Z in L^1 as well.

Define $\hat{Q}(H)$ to be the integral on the right side of 4.32. Convergence in L^1 allows us to write

$$\hat{Q}(H) = \mathbb{E}\ 1_H\ Z = \lim_n\ \mathbb{E}\ 1_H\ X_n$$

for every event H. But in view of 4.36, $\hat{Q}(H) = Q(H)$ for every H in $\mathcal{F}_n$; that is, $Q = \hat{Q}$ on $\mathcal{F}_n$ for each n. Hence, Q and $\hat{Q}$ coincide on the p-system $\cup_n \mathcal{F}_n$ and, therefore, on the σ-algebra $\mathcal{G}$ generated by that p-system; see Proposition 3.8 of Chapter I. □

4.41 Remark. *Singularity.* Suppose that Q on $(\Omega, \mathcal{G})$ is singular with respect to P. Lemma 4.37 still holds, the almost sure limit Z is positive and integrable, and the right side of 4.32 defines a finite measure $\hat{Q}$. Using Fatou's lemma with 4.36 shows that $\hat{Q}(H) \leq Q(H)$, which means that $\hat{Q}$ puts all its mass on a set of zero P-measure. But, obviously, $\hat{Q}$ is absolutely continuous with respect to P. It follows that $\hat{Q} = 0$, which means that $Z = 0$ almost surely.

4.42 Remark. *Lebesgue's decomposition.* Let the measure Q be an arbitrary finite measure on $(\Omega, \mathcal{G})$. We may normalize it to make it a probability measure, and we assume so. Then, $\hat{P} = {}^1/_2(P + Q)$ is a probability measure, and both P and Q are absolutely continuous with respect to $\hat{P}$. Thus, by the preceding theorem, there exists a positive X in $\mathcal{G}$ such that

$$P(H) = \int_H \hat{P}(d\omega) X(\omega), \quad Q(H) = \int_H \hat{P}(d\omega)\,(2 - X(\omega))$$

and we may assume that $0 \leq X \leq 2$. Thus, with $Z = \frac{2}{X} - 1$ and $\Omega_0 = \{X = 0\}$,

4.43 $$Q(H) = \int_H P(d\omega)\, Z(\omega) + Q(H \cap \Omega_0), \quad H \in \mathcal{G}.$$

On the right side, the integral defines a measure Q_c which is absolutely continuous with respect to P, and the second term defines a measure Q_s which is singular with respect to P. This decomposition $Q = Q_c + Q_s$ is called the Lebesgue decomposition of Q.

Doob's theorem for families of measures

This is immensely useful in the theory of Markov processes. The separability condition cannot be removed. In applications, E is sometimes a "space", sometimes the time set, and sometimes is the space-time product. The point of the theorem is the joint measurability of Z.

4.44 THEOREM. *Let Ω be a set and $\mathcal{G}$ a separable σ-algebra on it. Let $(E, \mathcal{E})$ be an arbitrary measurable space. Let Q be a bounded transition kernel, and P a probability kernel, both from $(E, \mathcal{E})$ into $(\Omega, \mathcal{G})$. Suppose that, for each x in E, the measure $H \mapsto Q(x, H)$ is absolutely continuous with respect to the measure $H \mapsto P(x, H)$. Then, there exists a positive Z in $\mathcal{E} \otimes \mathcal{G}$ such that*

4.45 $$Q(x, H) = \int_H P(x, d\omega) Z(x, \omega), \quad x \in E, H \in \mathcal{G}.$$

Proof. For each x, define $X_n(x, \omega)$ by 4.35 from the measures $Q(x, \cdot)$ and $P(x, \cdot)$. Measurability of $Q(x, H)$ and $P(x, H)$ in x shows that, for each r in $\mathbb{R}_+$,

$$\{(x, \omega) \in E \times \Omega : \quad X_n(x, \omega) \le r\}$$

is a finite union of rectangles $A \times H$ in $\mathcal{E} \otimes \mathcal{F}_n$. Thus, all the X_n are in $\mathcal{E} \otimes \mathcal{G}$ (this is joint measurability), which implies that Z defined next is in $\mathcal{E} \otimes \mathcal{G}$: For x in E and ω in Ω, let

$$Z(x, \omega) = \lim_n X_n(x, \omega)$$

if the limit exists and is finite, and otherwise, put $Z(x, \omega) = 0$. Now, for each x, Theorem 4.31 shows that 4.45 holds. □

Exercises and complements

4.46 *Doob's decomposition.* Let X be an L^1-bounded submartingale, and let $X = X_0 + M + A$ be its Doob decomposition. Show that the martingale M and the increasing process A are both convergent and their limits M_∞ and A_∞ are integrable.

4.47 *Convergence in L^p.* If $Z \in L^p$ for some p in $[1, \infty]$, then the martingale $(\mathbb{E}_n Z)$ is L^p-bounded and converges to $\mathbb{E}_\infty Z$ almost surely and in L^p.

4.48 *Dominated convergence in L^p.* Let X be a martingale that is L^p-bounded for some $p > 1$. Define

$$X^* = \sup_n |X_n|.$$

a) Show that $X^* \in L^p$ and that (X_n) is dominated by X^*.

b) Show that X converges almost surely and in L^p to a random variable X_∞, and $|X_\infty| \le X^*$.

4.49 *Convergence in reversed time.* In Theorem 4.17, suppose further that $X_0 \in L^p$ for some p in $[1, \infty]$. Show that, then, X converges to $X_{-\infty}$ in L^pas well. In particular, show that $X^* = \sup_n |X_n|$ is in L^p.

4.50 *Markov chains started at $-\infty$.* Let $X = (X_n)_{n \in \mathbb{T}}$ be a Markov chain with state space $(E, \mathcal{E})$ and transition kernel P over the time-set $\mathbb{T} = \{\ldots, -2, -1, 0\}$; see Definition 2.11 and take $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{T}}$ and require 2.12 for n in $\mathbb{T}$. Let f be a bounded function in $\mathcal{E}_+$, and set

$$M_n f = (P^{-n} f) \circ X_n, \quad n \in \mathbb{T}.$$

a) Show that $M_n f = \mathbb{E}_n\, f \circ X_0$ for $n \in \mathbb{T}$. Show that it converges, as $n \to -\infty$, to a bounded random variable mf almost surely and in L^1.

b) Suppose that there is a probability measure μ on $(E, \mathcal{E})$ such that $P^n f \to \mu f$ as $n \to +\infty$ for every bounded f in $\mathcal{E}_+$. Show that, then, the random variable mf is equal to the constant μf, the integral of f with respect to the measure μ. Thus, we have shown that

$$\lim_{n \to -\infty} \mathbb{E}_n\, f \circ X_0 = \mu f$$

for every bounded f in $\mathcal{E}_+$, which enables us to interpret μ as "the distribution of X_0 assuming that the chain is started at $-\infty$".

4.51 *Submartingales majorized by martingales.* Let $X = (X_n)_{n \in \mathbb{N}}$ be a uniformly integrable submartingale and let $X_\infty = \lim X_n$. Define

$$M_n = \mathbb{E}_n\, X_\infty, \quad n \in \mathbb{N},$$

Then, M is a uniformly integrable martingale. Show that $X_n \le M_n$ almost surely for all n.

4.52 *Decomposition of submartingales.* Let X be an L^1-bounded submartingale and let $X_\infty = \lim X_n$. Define

$$M_n = \mathbb{E}_n X_\infty, \quad V_n = X_n - M_n, \quad n \in \mathbb{N}.$$

This yields a decomposition $X = M + V$, where M is a uniformly integrable martingale and V is a submartingale with $\lim V_n = 0$ almost surely. Show these, and show that this decomposition is unique.

4.53 *Potentials.* Let X be a positive supermartingale. (Then, it is convergent almost surely.) If $\lim X_n = 0$ almost surely, then X is called a *potential.* Show that a positive supermartingale X is a potential if $\lim \mathbb{E}\, X_n = 0$.

4.54 *Decomposition of supermartingales.* Let X be a supermartingale with $\sup \mathbb{E}\, X_n^- < \infty$. Show that, then, X converges almost surely to an integrable random variable X_∞. Show that there is a unique decomposition

$$X = M + V$$

where M is a uniformly integrable martingale and V is a potential.

4.55 *Riesz decomposition.* Every positive supermartingale X has a decomposition

$$X = Y + Z$$

where Y is a positive martingale and Z is a potential. This is called the Riesz decomposition of X. Show this by following the steps below:

a) Show that the limit $Y_m = \lim_{n\to\infty} \mathbb{E}_m X_{m+n}$ exists almost surely.

b) Show that $Y = (Y_m)$ is a positive martingale.

c) Show that $Z = X - Y$ is a positive supermartingale and use 4.53 to conclude that Z is a potential.

4.56 *Continuation.* The martingale Y in the preceding decomposition is the maximal submartingale majorized by X, that is, if W is a submartingale and $W_n \le X_n$ for every n then $W_n \le Y_n$ for every n.

4.57 *Another decomposition for supermartingales.* Let X be a supermartingale with $\sup \mathbb{E} X_n^- < \infty$. Write $X = M + V$ as in Exercise 4.54. Let $V = N + Z$ be the Riesz decomposition of V. Then,

$$X = M + N + Z,$$

where M is a uniformly integrable martingale, N is a martingale potential, and Z is a potential.

4.58 *Krickeberg decomposition.* Let X be an L^1-bounded martingale. Then $X^+ = (X_n^+)$ and $X^- = (X_n^-)$ are positive submartingales. Show that

$$Y_n = \lim_m \mathbb{E}_n\, X_{n+m}^+, \quad Z_n = \lim_m \mathbb{E}_n\, X_{n+m}^-$$

exist. Show that Y and Z are positive and L^1-bounded martingales. Show that

$$X = Y - Z.$$

This is called the Krickeberg decomposition.

4.59 *Continuation.* A martingale is L^1-bounded if and only if it is the difference of two positive L^1-bounded martingales. Show.

4.60 *Continuation.* In the Krickeberg decomposition of an L^1-bounded martingale X, the process Y is the minimal positive martingale majorizing X, and the process Z is the minimal positive martingale majorizing $-X$ (see 4.56 for the meaning). Show this.

5 Martingales in Continuous Time

Throughout this section, $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space in the background, and $\mathcal{F}$ is a filtration over $\mathbb{R}_+$ which is extended onto $\bar{\mathbb{R}}_+$ as usual by setting $\mathcal{F}_\infty = \lim \mathcal{F}_t = \vee_t \mathcal{F}_t$. We shall assume that $\mathcal{F}$ satisfies two technical conditions. First, we assume that $\mathcal{F}$ is *right-continuous*, that is,

$$\mathcal{F}_t = \cap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}, \quad t \in \mathbb{R}_+.$$

Heuristically, this means that $\mathcal{F}_t$ includes all events that can be told by an "infinitesimal peek" into the future. Second, we assume that $\mathcal{F}$ is *augmented*, which means that $(\Omega, \mathcal{H}, \mathbb{P})$ is complete and that $\mathcal{F}_0$ (and therefore every $\mathcal{F}_t$) contains the collection of all negligible events in $\mathcal{H}$. These two conditions are considered harmless and are generally easy to fulfill; we shall clarify these concepts in Section 7 along with their ramifications.

Throughout, X is a real-valued stochastic process indexed by $\mathbb{R}_+$ or $\bar{\mathbb{R}}_+$ and adapted to $\mathcal{F}$. We shall assume that it is right-continuous and has left-limits on $\mathbb{R}_+$. We shall show in Section 7 that all martingales can be modified to have such regularity properties; thus, these assumptions on X are harmless.

All martingales, stopping times, etc. will be relative to the filtration $\mathcal{F}$ unless stated otherwise. As always, Convention 1.21 is in force: $\mathbb{E}_T$ denotes the conditional expectation operator given $\mathcal{F}_T$. Since X is adapted and right-continuous, Theorem 1.14 characterizing $\mathcal{F}_T$ shows that X_T belongs to $\mathcal{F}_T$, but one should take care that X_T be well-defined (for ω with $T(\omega) = +\infty$).

Doob martingales

For martingales in discrete time, Doob's stopping theorem 3.11 extends the martingale property at fixed times to random times that are bounded stopping times; see also Theorem 4.12 and Remark 4.13b. The resulting "strong" martingale equality (3.12 with the equality sign) captures the essence of martingales. We isolate this and incorporate it into the following definitions.

5.1 Definition. *The process X is said to have the* Doob property for *(S, T) provided that S and T be stopping times with $S \leq T$, and X_S and X_T be well-defined and integrable, and*

$$X_S = \mathbb{E}_S\ X_T.$$

5.2 Definition. *The process X is said to be a* Doob martingale *on $[0, \zeta]$ if ζ is a stopping time and X has the Doob property for (S, T) for all stopping times S and T with $0 \leq S \leq T \leq \zeta$.*

5.3 Remarks. a) For the Doob property, the condition that X_S and X_T be well-defined is needed only when X is indexed by $\mathbb{R}_+$ and S or T is allowed to take $+\infty$ as a value. If X is a Doob martingale on $[0, \zeta]$, there is the implicit assumption that X_ζ is well-defined and integrable and has expectation equal to $\mathbb{E}\ X_0$; these follow from the assumed Doob property for (S, T) with $S = 0$ and $T = \zeta$.

b) Note that Doob martingales are defined for closed intervals $[0, \zeta]$. Being closed on the right plays a significant role in the treatment below.

c) Suppose that X is a Doob martingale on $[0, \zeta]$. Then, the Doob property for $(t \wedge \zeta, \zeta)$ implies that

$$\hat{X}_t = X_{\zeta \wedge t} = \mathbb{E}_{\zeta \wedge t} X_\zeta = \mathbb{E}_t \mathbb{E}_\zeta X_\zeta = \mathbb{E}_t X_\zeta$$

for every t in $\mathbb{R}_+$. Thus, $(\hat{X}_t)_{t\in\mathbb{R}_+}$ is a uniformly integrable martingale; see Proposition 2.7. In other words, if X is a Doob martingale on $[0,\zeta]$ then the process $\hat{X}$ obtained by stopping X at ζ is a uniformly integrable martingale. We shall show below, in Theorem 5.14, that the converse is true as well.

In the following, we characterize Doob martingales in terms of simpler looking conditions, show their intimate connections to uniform integrability, and discuss some of their uses on Brownian motions. We start with the following characterization; see Remark 3.15 for the discrete-time source of the ideas.

5.4 Theorem. *Let ζ be a stopping time. Then, the following are equivalent:*

a) The process X is a Doob martingale on $[0,\zeta]$.
b) For every stopping time T majorized by ζ,

$$X_T = \mathbb{E}_T X_\zeta.$$

c) For every stopping time T majorized by ζ,

$$\mathbb{E}\ X_T = \mathbb{E}\ X_0.$$

Proof. Clearly (a)⇒(b): the latter is the Doob property for (T,ζ). If (b) holds, then $\mathbb{E}\ X_T = \mathbb{E}\ X_\zeta$ and taking $T=0$ we get $\mathbb{E}\ X_\zeta = \mathbb{E}\ X_0$; so, (b)⇒(c).

To show that (c)⇒(b), assume (c). Let T be a stopping time majorized by ζ, that is, $T \le \zeta$. Take an event H in $\mathcal{F}_T$ and define

$$S = T\ 1_H + \zeta(1 - 1_H).$$

Then, S is a stopping time majorized by ζ, and

$$X_\zeta - X_S = (X_\zeta - X_T)\ 1_H.$$

The expectation of the left side is 0 since $\mathbb{E}\ X_\zeta = \mathbb{E}\ X_0 = \mathbb{E}\ X_S$ by the assumed property (c). Thus, the expectation of the right side is 0, and this is for arbitrary H in $\mathcal{F}_T$; hence, (b) holds.

Finally, (b)⇒(a): If S and T are stopping times with $S \le T \le \zeta$, then

$$\mathbb{E}_S X_T = \mathbb{E}_S \mathbb{E}_T X_\zeta = \mathbb{E}_S X_\zeta = X_S,$$

where we used (b) to justify the first and the last equalities; this shows that Doob property holds for (S,T). □

5.5 Corollary. *If X is a Doob martingale on $[0,\zeta]$, then*

$$\{X_T :\ T \text{ is a stopping time, } T \le \zeta\} = \{X_{T\wedge\zeta} :\ T \text{ is a stopping time}\}$$

is uniformly integrable.

Proof. The proof is immediate from the statement (b) of the preceding theorem and Lemma 2.8. □

Doob's stopping theorem

The statement (c) of the next theorem is the classical version of Doob's stopping theorem for martingales served up in the language of Definition 5.1.

5.6 THEOREM. *The following are equivalent for X:*

a) It is a martingale on $\mathbb{R}_+$.
b) It is a Doob martingale on $[0, b]$ for every b in $\mathbb{R}_+$.
c) It has the Doob property for (S, T) whenever S and T are bounded stopping times with $S \leq T$.

Proof. All the implications are immediate from the definitions except for the implication (a)⇒(b). To show it, we use the preceding characterization theorem. Accordingly, assume (a), fix b in $\mathbb{R}_+$, and let T be a stopping time bounded by b. Then, b is a fixed stopping time, X_T is well-defined, and we are to show that $\mathbb{E}\ X_T = \mathbb{E}\ X_0$.

Let (T_n) be as in Proposition 1.20. For fixed n, the stopping time T_n is bounded by $b + 1$ and takes values in the discrete set $\mathbb{T}$ consisting of the numbers $b+1$ and $k/2^n$ with k in $\mathbb{N}$. By Doob's stopping theorem 3.11 for the discrete-time martingale $(X_t)_{t\in\mathbb{T}}$ applied at the bounded times T_n and $b+1$,

5.7 $$X_{T_n} = \mathbb{E}_{T_n} X_{b+1}.$$

Recalling that $\cdots \leq T_2 \leq T_1 \leq T_0$, this means that $(X_{T_n})_{n\in\mathbb{N}}$ is a reversed-time martingale relative to the filtration $(\mathcal{F}_{T_n})$. Thus, by Theorem 4.17, it converges almost surely and in L^1 to an integrable random variable. But, since (T_n) decreases to T and X is right-continuous, that limit is X_T. Since convergence in L^1 implies the convergence of expectations, and in view of 5.7,

$$\mathbb{E}\ X_T = \lim\ \mathbb{E}\ X_{T_n} = \mathbb{E}\ X_{b+1} = \mathbb{E}\ X_0. \qquad \square$$

The preceding theorem shows, in particular, that X is a Doob martingale on $[0, b]$ if and only if it is a martingale on $[0, b]$: in the proof, replace T_n by $T_n \wedge b$ and replace $\mathbb{T}$ with $\mathbb{T} \wedge b = \{t \wedge b : t \in \mathbb{T}\}$. This remains true when b is replaced by $+\infty$, as the next theorem shows. Note that the second assertion is Doob's stopping theorem, for this case, extended to arbitrary stopping times. See 4.12 for the discrete-time version.

5.8 THEOREM. *The process X is a Doob martingale on $\bar{\mathbb{R}}_+$ if and only if it is a martingale on $\bar{\mathbb{R}}_+$. If so, then*

$$\mathbb{E}_S X_T = X_{S\wedge T}$$

for arbitrary stopping times S and T.

Proof. Assuming that X is a martingale on $\bar{\mathbb{R}}_+$, we shall show that

5.9 $$X_T = \mathbb{E}_T X_\infty$$

for every stopping time T. This will prove the first claim via Theorem 5.4 with $\zeta = +\infty$. The second claim follows from 5.9 since $\mathbb{E}_S\mathbb{E}_T = \mathbb{E}_{S\wedge T}$. So, assume X is a martingale on $\bar{\mathbb{R}}_+$ and let T be a stopping time.

For each n in $\mathbb{N}$, Theorem 5.4 implies that X has the Doob property for the bounded stopping times $T \wedge n$ and n:

$$X_{T\wedge n} = \mathbb{E}_{T\wedge n} X_n.$$

Since X is a martingale on $\bar{\mathbb{R}}_+$, the process $(X_n)_{n\in\bar{\mathbb{N}}}$ is a martingale relative to $(\mathcal{F}_n)_{n\in\bar{\mathbb{N}}}$; thus, $X_n = \mathbb{E}_n X_\infty$ for each n. This implies, together with the observations $\mathbb{E}_{T\wedge n}\mathbb{E}_n = \mathbb{E}_{T\wedge n} = \mathbb{E}_n\mathbb{E}_T$, that

5.10 $$X_{T\wedge n} = \mathbb{E}_n\mathbb{E}_T X_\infty, \quad n \in \mathbb{N}.$$

At this point, we remark that

5.11 $$\hat{\mathcal{F}}_\infty = \vee_{n\in\mathbb{N}}\mathcal{F}_n = \vee_{t\in\mathbb{R}_+}\mathcal{F}_t = \mathcal{F}_\infty.$$

This is because every $\mathcal{F}_n$ is contained in some $\mathcal{F}_t$ and vice versa.

The right side of 5.10 converges almost surely, by Corollary 4.11, to the conditional expectation of $\mathbb{E}_T X_\infty$ given $\hat{\mathcal{F}}_\infty$, which is the same as $\mathbb{E}_\infty\mathbb{E}_T X_\infty = \mathbb{E}_T X_\infty$ in view of 5.11. Whereas, the left side of 5.10 converges almost surely to X_T: if $T(\omega) < \infty$ then $X_{T\wedge n}(\omega) = X_T(\omega)$ for all n large enough, and if $T(\omega) = \infty$ then $X_{T\wedge n}(\omega) = X_n(\omega)$ for every n, which converges to $X_\infty(\omega)$ for almost every ω by the fact that $(X_n)_{n\in\bar{\mathbb{N}}}$ is a martingale (see Theorem 4.7). Hence, 5.9 holds. □

Uniform integrability

The best one can ask of a martingale is that it be a Doob martingale on $\bar{\mathbb{R}}_+$. Often, however, one starts with a martingale on $\mathbb{R}_+$.

5.12 THEOREM. *Suppose that X is a martingale on $\mathbb{R}_+$. Then, it can be extended to a Doob martingale on $\bar{\mathbb{R}}_+$ if and only if it is uniformly integrable.*

Proof. Suppose that the martingale can be extended to a Doob martingale $\bar{X}$ on $\bar{\mathbb{R}}_+$, that is, there exists a random variable X_∞ in $\mathcal{F}_\infty$ such that $\bar{X} = (X_t)_{t\in\bar{\mathbb{R}}_+}$ is a Doob martingale on $[0, +\infty]$. Then, Corollary 5.5 implies that X is uniformly integrable.

Conversely, suppose that X is a uniformly integrable martingale on $\mathbb{R}_+$. Then, in particular, $(X_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale. By Theorem 4.7, it converges almost surely and in L^1 to an integrable random variable X_∞ in $\mathcal{F}_\infty$ (see 5.11 to the effect that $\mathcal{F}_\infty$ is the limit of $(\mathcal{F}_n)_{n\in\mathbb{N}}$ as well as of $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$), and $X_n = \mathbb{E}_n X_\infty$ for every n. For t in $\mathbb{R}_+$, choose n in $\mathbb{N}$ so that $t < n$; then, since X is a martingale on $\mathbb{R}_+$,

$$X_t = \mathbb{E}_t X_n = \mathbb{E}_t\mathbb{E}_n X_\infty = \mathbb{E}_t X_\infty.$$

This shows that $\bar{X} = (X_t)_{t\in\bar{\mathbb{R}}_+}$ is a martingale and, therefore, is a Doob Martingale on $\bar{\mathbb{R}}_+$ in view of Theorem 5.8 above. □

Together with Proposition 2.7, the preceding proof shows the following as well.

5.13 THEOREM. *The process X is a uniformly integrable martingale on $\mathbb{R}_+$ if and only if*

$$X_t = \mathbb{E}_t Z, \quad t \in \mathbb{R}_+,$$

for some integrable random variable Z. Moreover, then, $X_\infty = \lim X_t$ exists almost surely and in L^1 and satisfies $X_\infty = \mathbb{E}_\infty Z$, and $\bar{X} = (X_t)_{t \in \bar{\mathbb{R}}_+}$ is a Doob martingale on $\bar{\mathbb{R}}_+$.

Stopped martingales

5.14 THEOREM. *Let ζ be a stopping time. Let $\hat{X}$ be the process X stopped at ζ, that is, $\hat{X}_t = X_{t \wedge \zeta}$ for t in $\mathbb{R}_+$.*

a) If X is a martingale, $\hat{X}$ is a martingale.

b) The process X is a Doob martingale on $[0, \zeta]$ if and only if $\hat{X}$ is a uniformly integrable martingale.

Proof. a) Suppose that X is a martingale on $\mathbb{R}_+$. Let T be a stopping time bounded by some b in $\mathbb{R}_+$. Then, $T \wedge \zeta$ is a stopping time bounded by b, and X is a Doob martingale on $[0, b]$ by Theorem 5.6, which together yield $\mathbb{E}\ X_{T \wedge \zeta} = \mathbb{E}\ X_0$ via Theorem 5.4. But, $X_{T \wedge \zeta} = \hat{X}_T$ and $X_0 = \hat{X}_0$. So,

$$\mathbb{E}\ \hat{X}_T = \mathbb{E}\ \hat{X}_0$$

for every stopping time bounded by some $b < \infty$, which implies via Theorem 5.4 that $\hat{X}$ is a Doob martingale on $[0, b]$ for every b in $\mathbb{R}_+$, which in turn implies via Theorem 5.6 that $\hat{X}$ is a martingale.

b) Necessity part of the statement (b) was shown in Remark 5.3c. To show the sufficiency part, suppose that $\hat{X}$ is a uniformly integrable martingale on $\mathbb{R}_+$. By Theorem 5.13, we can extend it to a Doob martingale on $\bar{\mathbb{R}}_+$ by defining $\hat{X}_\infty = \lim \hat{X}_t$. Then, for every stopping time T majorized by ζ, we have $X_T = \hat{X}_T$ and

$$\mathbb{E}\ X_T = \mathbb{E}\ \hat{X}_T = \mathbb{E}\ \hat{X}_0 = \mathbb{E}\ X_0$$

by Theorem 5.4 for $\hat{X}$. Thus, by 5.4 again, X is a Doob martingale on $[0, \zeta]$. □

Criteria for being Doob

The following criterion is easy to fulfill in many applications.

5.15 PROPOSITION. *Suppose that X is a martingale on $\mathbb{R}_+$. Let ζ be a stopping time. Suppose that X is dominated on $[0, \zeta] \cap \mathbb{R}_+$ by an integrable random variable. Then, the almost sure limit $X_\zeta = \lim_{t \to \infty} X_{\zeta \wedge t}$ exists and is integrable, and X is a Doob martingale on $[0, \zeta]$.*

Proof. Let X and ζ be such. Let Z be an integrable random variable such that, for almost every ω, $|X_t(\omega)| \le Z(\omega)$ for every t in $\mathbb{R}_+$ with $t \le \zeta(\omega)$. Define $\hat{X}$ to be X stopped at ζ.

By Theorem 5.14a, then, $\hat{X}$ is a martingale on $\mathbb{R}_+$. By assumption, $\hat{X}$ is dominated by the integrable random variable Z almost surely, which implies that $\hat{X}$ is uniformly integrable. Thus, the almost sure (and in L^1) limit

$$\lim_{t\to\infty} \hat{X}_t = \lim_{t\to\infty} X_{\zeta\wedge t} = X_\zeta$$

exists and is integrable. It follows from Theorem 5.14b that X is a Doob martingale on $[0, \zeta]$. □

5.16 Example. Let X be a continuous martingale. For fixed integer $n \ge 1$, let

$$\zeta_n = \inf\{t \ge 0 : |X_t| \ge n\}.$$

Then, ζ_n is a stopping time, and X is dominated by the constant n on $[0, \zeta_n] \cap \mathbb{R}_+$. The preceding theorem implies that X is a Doob martingale on $[0, \zeta_n]$.

Local martingales and semimartingales

The modern theory of stochastic analysis is built around these objects. Our aim is to provide a bridge to it by introducing the terms.

5.17 Definition. *Let ζ be a stopping time. The process X is called a* local martingale *on* $[0, \zeta)$ *if there exists an increasing sequence of stopping times ζ_n with limit ζ such that $(X_t - X_0)_{t\in\mathbb{R}_+}$ is a Doob martingale on $[0, \zeta_n]$ for every n. If it is a local martingale on $\mathbb{R}_+ = [0, \infty)$, then it is simply called a local martingale.*

Every martingale is a local martingale, because, if X is a martingale, then it is a Doob martingale on $[0, n]$ and the definition is satisfied with $\zeta_n = n$ and $\zeta = +\infty$. Of course, if X is a Doob martingale on $[0, \zeta]$, then it is a local martingale on $[0, \zeta)$ trivially (take $\zeta_n = \zeta$ for all n).

In the definition above, the sequence (ζ_n) is called a *localizing sequence.* In general, there are many localizing sequences for the same local martingale. Choosing the correct one is an art and depends on the application at hand. For example, if X is a continuous martingale as in Example 5.16, one localizing sequence is given by $\zeta_n = n$, another by the ζ_n defined there; the latter has the advantage of making X a bounded (Doob) martingale on $[0, \zeta_n]$ for each n. In general, it is worth noting that, if (ζ'_n) and (ζ''_n) are localizing sequences, then so is $(\zeta'_n \wedge \zeta''_n)$.

Localization is used in other contexts as well. For instance, a process (V_t) is said to be *locally of finite variation* if there exists a sequence (ζ_n) of stopping times increasing to $+\infty$ almost surely such that, for every ω, the path $t \mapsto V_t(\omega)$ has finite variation over the interval $[0, \zeta_n(\omega)]$ for every n.

5.18 DEFINITION. *The process X is called a semimartingale if it can be decomposed as*

$$X = L + V$$

where L is a local martingale and V is locally of finite variation.

In the definition, it is understood that both L and V are to be adapted to the same filtration $\mathcal{F}$ to which X is adapted. The localizing sequences for L and V can differ, but it is always possible to find a sequence that localizes both.

Applications to Brownian motion

Our aim is to illustrate some uses of the foregoing with a sustained example or two. Many other problems can be solved by similar techniques. We start with the more delicate problem of showing that most hitting times of Brownian motion are almost surely finite.

Let $W = (W_t)_{t\in\mathbb{R}_+}$ be a Wiener process with respect to the filtration $\mathcal{F}$ on $\mathbb{R}_+$; see Definition 2.15. Define

5.19 $$T_a = \inf\{t > 0 : W_t \geq a\}, \quad a > 0,$$

the time W enters $[a, \infty)$ for the first time. It is a stopping time for each a.

5.20 PROPOSITION. *For each a in $(0, \infty)$ the stopping time T_a is almost surely finite, its expected value is $+\infty$, and its distribution and the corresponding Laplace transform are as follows:*

$$\mathbb{P}\{T_a \in B\} = \int_B dt \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}, \;\; B \in \mathcal{B}_{\mathbb{R}_+}; \quad \mathbb{E}\; e^{-pT_a} = e^{-a\sqrt{2p}}, \;\; p \in \mathbb{R}_+.$$

Proof. Fix a, write T for T_a. Fix $p > 0$, put $r = \sqrt{2p}$, and note that $p = r^2/2$. It was shown in Proposition 2.17 that

5.21 $$X_t = \exp(rW_t - pt), \quad t \in \mathbb{R}_+,$$

defines a martingale. For arbitrary ω, if $T(\omega) < \infty$ and $t \leq T(\omega)$, then $W_t(\omega) \leq a$ by 5.19 and the continuity of W, which implies that $X_t(\omega) \leq e^{ra}$, and the same holds for $t < \infty$ when $T(\omega) = +\infty$. In other words, X is bounded by the constant e^{ra} on $[0, T] \cap \mathbb{R}_+$. Thus, by Proposition 5.15, X_T is well-defined and integrable, and X is a Doob martingale on $[0, T]$. By Theorem 5.4, then,

5.22 $$\mathbb{E}\; X_T = \mathbb{E}\; X_0 = 1.$$

On the event $\{T = +\infty\}$, we have $W_t \leq a$ for all t, which implies that $X_T = \lim X_t = 0$. Whereas, on the event $\{T < \infty\}$, we have $W_T = a$ and $X_T = \exp(ra - pT)$. Hence,

$$e^{ra-pT} = e^{ra-pT} 1_{\{T<\infty\}} = X_T 1_{\{T<\infty\}} = X_T,$$

which yields, by 5.22,

$$\mathbb{E}\ e^{-pT} = e^{-ra} = e^{-a\sqrt{2p}}$$

(recall that $r = \sqrt{2p}$). Since $p > 0$ was arbitrary, this shows that the Laplace transform for T is as claimed. Letting $p \to 0$, we see that

$$\mathbb{P}\{T < \infty\} = \lim_{p \to 0} \mathbb{E}\ e^{-pT} = 1.$$

Inverting the Laplace transform yields the claimed distribution for T. Using the distribution to compute the expectation we get $\mathbb{E}\ T = +\infty$. □

5.23 COROLLARY. *Almost surely,*

$$T_0 = \inf\{t > 0 : W_t > 0\} = 0.$$

Proof. Clearly, $0 \le T_0 \le T_a$ for every $a > 0$. Thus, for $p > 0$,

$$\mathbb{E}\ e^{-pT_0} \ge \mathbb{E}\ e^{-pT_a} = e^{-a\sqrt{2p}}$$

for every $a > 0$. This shows that the left side is equal to 1, which implies in turn that $T_0 = 0$ almost surely. □

Similarly to 5.19, we define the entrance times

5.24 $$T_a = \inf\{t > 0 : W_t \le a\}, \quad a < 0.$$

Since $(-W_t)$ is again a Wiener process, T_a and T_{-a} have the same distribution for every a, and the distribution is given by 5.20 for $a > 0$. We state this next and add a remark or two whose proofs are left as exercises.

5.25 COROLLARY. *For every non-zero a in $\mathbb{R}$, the stopping time T_a is almost surely finite, has expected value $+\infty$, and has the same distribution as a^2/Z^2 where Z is a standard Gaussian variable.*

For $a > 0$ for instance, T_a is the amount of time W spends in the interval $(-\infty, a)$ before exiting it. The interval being unbounded, $\mathbb{E}\ T_a = +\infty$. Otherwise, W exits every bounded interval in finite time with finite expected value. We show this next along with related results. Define

5.26 $$T = \inf\{t : W_t \notin (a, b)\}, \quad a < 0 < b,$$

that is, T is the time of exit from (a, b); recall that $W_0 = 0$. Obviously,

5.27 $$T = T_a \wedge T_b,$$

and we have shown above that the entrance times T_a and T_b are almost surely finite. Thus, $T < \infty$ almost surely, and W_T is either a or b, with some

probabilities p_a and p_b respectively, $p_a + p_b = 1$. Since the martingale W is bounded on the time interval $[0,T]$, it is a Doob martingale on $[0,T]$. It follows that $\mathbb{E}\ W_T = \mathbb{E}\ W_0 = 0$; in other words, $ap_a + bp_b = 0$. So,

5.28 $$p_a = \mathbb{P}\{W_T = a\} = \frac{b}{b-a}, \quad p_b = \mathbb{P}\{W_T = b\} = \frac{-a}{b-a}.$$

In order to compute the expected value of the exit time T, we consider the martingale Y defined in 2.19, that is, $Y_t = W_t^2 - t$, $t \in \mathbb{R}_+$. By Theorem 5.6, it has the Doob property for the bounded stopping times 0 and $T \wedge t$, that is, $\mathbb{E}_0 Y_{T\wedge t} = Y_0 = 0$. Hence,

$$\mathbb{E}\ (T \wedge t) = \mathbb{E}\ (W_{T\wedge t})^2, \quad t \in \mathbb{R}_+.$$

Since $T \wedge t$ increases to T as $t \to \infty$, the left side increases to $\mathbb{E}\ T$ by the monotone convergence theorem. Since $(W_{T\wedge t})^2$ is bounded by $a^2 \vee b^2$, and converges to W_T^2, the right side goes to $\mathbb{E}\ W_T^2$ by the bounded convergence theorem. Hence,

5.29 $$\mathbb{E}\ T = \mathbb{E}\ W_T^2 = (-a) \cdot b, \quad a < 0 < b,$$

in view of 5.28. Incidentally, we have also shown that Y is a Doob martingale on $[0,T]$.

Finally, we specify the distribution of the time of exit from a symmetric interval by means of Laplace transforms.

5.30 Proposition. *Let T be the first time that W exits the interval $(-a,a)$, where $a > 0$ is a fixed constant. Then, $\mathbb{E}\ T = a^2$ and*

$$\mathbb{E}\ e^{-pT} = 2/\left(e^{a\sqrt{2p}} + e^{-a\sqrt{2p}}\right), \quad p \in \mathbb{R}_+.$$

Proof. Fix $p > 0$, put $r = \sqrt{2p}$, and let X be as in 5.21. Then, X is a positive martingale bounded by e^{ra} in $[0,T]$, and $T < \infty$ almost surely, and W_T is well-defined and bounded. So, X is a Doob martingale on $[0,T]$, and $\mathbb{E}\ X_T = \mathbb{E}\ X_0 = 1$ by 5.4, that is,

$$\mathbb{E}\ \exp(rW_T - pT) = 1.$$

Note that T is also the exit time from $(-a,a)$ by the process $(-W_T)$; this is because the interval is symmetric. Hence,

$$\mathbb{E}\ \exp(-rW_T - pT) = 1.$$

Adding the last two equations side by side, we get

$$\mathbb{E}\ [\exp(rW_T) + \exp(-rW_T)]\,[\exp(-pT)] = 2.$$

Whether W_T is equal to a or $-a$, the first factor inside the expectation is equal to $e^{ra} + e^{-ra}$, which constant can come out of the expectation. So,

$$(e^{ra} + e^{-ra})\mathbb{E}\ e^{-pT} = 2,$$

which yields the claimed Laplace transform once we put $r = \sqrt{2p}$. □

Exercises and Complements

5.31 *Stopped martingales.* Let S be a stopping time and let Y be the process X stopped at S.

a) If X is a Doob martingale on $[0, \zeta]$, then Y is a Doob martingale on $[0, \zeta]$. Show.

b) Use (a) to prove that if X is a martingale then Y is a martingale.

5.32 *Continuation.* If X is a local martingale on $[0, \zeta)$, then Y is a local martingale on $[0, \zeta)$. Show.

5.33 *Doob's maximal inequalities.* Let $X = (X_t)$ be a submartingale that is positive and continuous. Let

$$M_t = \max_{0 \le r \le t} X_s.$$

Show that, for $p \ge 1$ and $b \ge 0$,

$$b^p \, \mathbb{P}\{M_t > b\} \;\le\; \mathbb{E}\, X_t^p.$$

Show that, if $X_t \in L^p$ for some $p > 1$, then, with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}\, M_t^p \;\le\; q^p \, \mathbb{E}\, X_t^p.$$

Hint: Let $D_n = \{tk/2^n : k = 0, 1, \dots, 2^n\}$; note that M_t is the limit, as $n \to \infty$, of $\max_{s \in D_n} X_s$; Use the discrete time results for the latter maxima.

5.34 *Convergence theorem in continuous-time.* Let X be a right-continuous submartingale. Suppose that it is L^1-bounded, that is, $\sup_{t \in \mathbb{R}_+} \mathbb{E}\, |X_t| < \infty$, which condition is equivalent to having $\lim_{t\to\infty} \mathbb{E}\, X_t^+ < \infty$. Then, the almost sure limit $X_\infty = \lim_{t\to\infty} X_t$ exists and is integrable. If X is uniformly integrable, then the convergence is in L^1 as well and $\bar{X} = (X_t)_{t \in \bar{\mathbb{R}}_+}$ is a submartingale.

5.35 *Reverse-time convergence.* Let $X = (X_t)_{t>0}$ be a right-continuous submartingale. Suppose that $\sup_{t \le 1} \mathbb{E}|X_t| < \infty$.

a) Show that the condition is equivalent to $\lim_{t\to 0} \mathbb{E}\, X_t < \infty$.

b) Show that $\lim_{t \to 0} X_t$ exists almost surely and in L^1.

Supplements for Brownian motion

Throughout the following exercises, W is a Wiener process with respect to the filtration $\mathcal{F}$.

5.36 *Distribution of* T_a. Let T_a be as defined by 5.19. Show that T_a has the same distribution as a^2/Z^2, where Z is a standard Gaussian variable. Note that the same is true of T_a with $a < 0$ as well.

5.37 *Continuation.* Show that $\mathbb{P}\{T_a \le t\} = \mathbb{P}\{|W_t| > a\}$.

5.38 *Maxima and minima.* Define

$$M_t = \max_{s \le t} W_s, \quad m_t = \min_{s \le t} W_s.$$

a) Show that $M_t(\omega) \ge a$ if and only if $T_a(\omega) \le t$, this being true for all $a > 0$, $t > 0$, and ω in Ω.

b) Show that M_t has the same distribution as $|W_t|$, and m_t the same distribution as $-|W_t|$.

5.39 *Continuation.* Show that the following are true for almost every ω:

a) $T_a(\omega) < \infty$ for every a in $\mathbb{R}$,
b) $t \mapsto M_t(\omega)$ is continuous, real-valued, and increasing with limit $+\infty$,
c) $t \mapsto m_t(\omega)$ is continuous, real-valued, and decreasing with limit $-\infty$.
d) $\liminf W_t(\omega) = -\infty$, $\limsup W_t(\omega) = +\infty$.
e) The set $\{t \in \mathbb{R}_+ : W_t(\omega) = 0\}$ is unbounded, that is, for every $b < \infty$ there is $t > b$ such that $W_t(\omega) = 0$. Consequently, there is a sequence (t_n), depending on ω, such that $t_n \nearrow +\infty$ and $W_{t_n}(\omega) = 0$ for every n.

5.40 *Exponential martingale.* Let $X_t = \exp(rW_t - \frac{1}{2}r^2 t)$ where r is a fixed real number. Since X is a positive martingale, $X_\infty = \lim X_t$ exists almost surely. Identify the limit. Is X uniformly integrable?

5.41 *Brownian motion.* Let $B_t = at + bW_t$, $t \ge 0$, where a and b are fixed numbers. Then, B is called a Brownian motion with drift rate a and volatility b and with $B_0 = 0$. Suppose that $a > 0$, $b > 0$, and fix $x > 0$. Show that

$$T = \inf\{t : B_t \ge x\}$$

is finite almost surely. Use the exponential martingale with $p > 0$ and $r = -\frac{a}{b} + \sqrt{\frac{a^2}{b^2} + 2p}$ to get

$$\mathbb{E}\, e^{-pT} = e^{-xr/b}, \quad \mathbb{E}\, T = x/a, \quad \mathrm{Var}\, T_x = xb^2/a^3.$$

5.42 *Brownian motion with negative drift.* Let $a > 0$ and put $B_t = W_t - at$, $t \in \mathbb{R}_+$. For $x > 0$, let $T = \inf\{t : B_t \ge x\}$.

a) Show that

$$\mathbb{E}\, e^{-pT} = \exp(-xa - x\sqrt{a^2 + 2p}), \quad p > 0.$$

Conclude, in particular, that $\mathbb{P}\{T < \infty\} = e^{-2ax}$.

b) Let $M = \sup_{t \in \mathbb{R}_+} B_t$. Show that M has the exponential distribution with parameter $2a$.

5.43 *Exit from an interval.* With $a > 0$ and $b > 0$, put $B_t = at + bW_t$, $t \in \mathbb{R}_+$. For $x < 0 < y$, let

$$T = \inf\{t > 0 : B_t \notin (x, y)\}.$$

Show that T is almost surely finite. Compute the distribution of B_T. Compute the mean and variance of T.

5.44 *Multi-dimensional Wiener*. Let W be an n-dimensional Wiener process, that is, $W = (W^{(1)}, \ldots, W^{(n)})$ where the components are independent Wiener processes. Then, $R = |W|$ is called the *radial Brownian motion*, or *Bessel process* of index n; for v in $\mathbb{R}^n$ we write $|v|$ for the length of v. For fixed $r > 0$, let

$$T = \inf\{t : |W_t| \geq r\},$$

the time of exit from the open ball of radius r centered at the origin. Show that $T < \infty$ almost surely. Show that $\mathbb{E}\, T = r^2/n$.

5.45 *Behavior near zero*. Returning back to one-dimension, show that

$$T_0 = \inf\{t > 0 : W_t > 0\} = \inf\{t > 0 : W_t < 0\} = 0$$

almost surely. Show that the following are true for almost every ω:

a) For every $\varepsilon > 0$ there is u in $(0, \varepsilon)$ such that $W_u(\omega) > 0$ and there is s in $(0, \varepsilon)$ such that $W_s(\omega) < 0$.

b) There exist strictly positive sequences (s_n), (t_n), (u_n) depending on ω such that

$$u_1 > t_1 > s_1 > u_2 > t_2 > s_2 > \ldots, \quad \lim u_n = \lim t_n = \lim s_n = 0$$

and

$$W_{u_n}(\omega) > 0, \quad W_{t_n}(\omega) = 0, \quad W_{s_n}(\omega) < 0, \quad n \geq 1.$$

6 Martingale Characterizations for Wiener and Poisson

Our primary aim is to complete the proofs of Theorems 2.19 and 2.23, the characterizations of Wiener and Poisson processes in terms of martingales. We start with the Poisson case, because the needed preliminaries are of independent interest.

In this section, $\mathcal{F} = (\mathcal{F}_t)$ is a filtration on $\mathbb{R}_+$, not necessarily augmented or right-continuous. All processes are indexed by $\mathbb{R}_+$ and adapted to $\mathcal{F}$, all with state space $\mathbb{R}$. Considering a process $F = (F_t)$, we shall generally think of it as the mapping $(\omega, t) \mapsto F_t(\omega)$ from $\Omega \times \mathbb{R}_+$ into $\mathbb{R}$, and we may use the phrase "the process F on $\Omega \times \mathbb{R}_+$" to indicate that thought.

Predictability

This is the continuous-time analog of the concept introduced by Definition 3.1. We shall develop it briefly, just enough to serve our present needs.

6.1 Definition. *The σ-algebra on $\Omega \times \mathbb{R}_+$ generated by the collection*

$$\mathcal{F}^{pp} = \{H \times (s, t] :\ 0 \leq s < t < \infty, H \in \mathcal{F}_s\} \cup \{H \times \{0\} :\ H \in \mathcal{F}_0\}$$

is called the $\mathcal{F}$-predictable σ-algebra *and is denoted by* $\mathcal{F}^p$. *A stochastic process* $F = (F_t)$ *is said to be* $\mathcal{F}$-predictable *if the mapping* $(\omega, t) \mapsto F_t(\omega)$ *is* $\mathcal{F}^p$*-measurable.*

It is usual to simply say predictable instead of $\mathcal{F}$-predictable when there can be no confusion over the filtration involved, which is our present situation. The elements of $\mathcal{F}^{pp}$ are called the primitive sets, and their indicators are called *primitives*. The following proposition implies, in particular, that every adapted left-continuous process is predictable. In the converse direction, every predictable process is adapted, but might fail to be left-continuous.

6.2 PROPOSITION. *The predictable σ-algebra $\mathcal{F}^p$ is also generated by the collection $\mathcal{G}$ of all adapted left-continuous processes on $\Omega \times \mathbb{R}_+$.*

Proof. Every primitive process is adapted and left-continuous; thus, $\mathcal{F}^p \subset \sigma\mathcal{G}$. To show the converse, that $\sigma\mathcal{G} \subset \mathcal{F}^p$, we need to show that every G in $\mathcal{G}$ is predictable. Let G be in $\mathcal{G}$. Then, for every time t and outcome ω, the value $G_t(\omega)$ is the limit, as $n \to \infty$, of

$$G_t^{(n)}(\omega) = G_0(\omega)1_{\{0\}}(t) + \sum G_a(\omega)1_{(a,b]}(t),$$

where the sum is over all intervals $(a, b]$ with $a = k/2^n$ and $b = (k+1)/2^n$, $k \in \mathbb{N}$; this is by the left-continuity of $t \mapsto G_t(\omega)$. Thus, to show that G is predictable, it is enough to show that each $G^{(n)}$ is predictable, which in turn reduces to showing that every term of $G^{(n)}$ is predictable. But, for fixed $(a, b]$, the process $(\omega, t) \mapsto G_a(\omega)1_{(a,b]}(t)$ is clearly predictable, since $G_a \in \mathcal{F}_a$ by the adaptedness of G, and $\mathcal{F}^p$ is generated by the primitive processes; similarly, the process $(\omega, t) \mapsto G_0(\omega)1_{\{0\}}(t)$ is predictable. □

Martingales associated with some increasing processes

Let $N = (N_t)$ be an increasing right-continuous process adapted to the filtration $\mathcal{F}$. Let $\nu_t = \mathbb{E}\, N_t$ be finite for every t, and suppose that

6.3 $$\mathbb{E}_s(N_t - N_s) = \nu_t - \nu_s, \quad 0 \le s < t < \infty,$$

where $\mathbb{E}_s$ denotes the conditional expectation given $\mathcal{F}_s$ as usual. This is equivalent to assuming that

6.4 $$\tilde{N}_t = N_t - \nu_t, \quad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale. In particular, these conditions are fulfilled when N has independent increments and $\mathbb{E}\, N_t < \infty$.

6.5 THEOREM. *For every positive predictable process F,*

6.6 $$\mathbb{E} \int_{\mathbb{R}_+} F_t\, dN_t = \mathbb{E} \int_{\mathbb{R}_+} F_t\, d\nu_t.$$

Remark. The integrals above are Stieltjes integrals; for instance, the one on the left defines a random variable V where $V(\omega)$ is the integral of $t \mapsto F_t(\omega)$ with respect to the measure on $\mathbb{R}_+$ defined by the increasing right-continuous function $t \mapsto N_t(\omega)$.

Proof. Consider the collection of all positive predictable processes F for which 6.6 holds. That collection is a monotone class: it includes the constants, it is a linear space, and it is closed under increasing limits, the last being the result of the monotone convergence theorem applied twice on the left side of 6.6 and twice on the right. Thus, the monotone class theorem will conclude the proof once we show that 6.6 is true for primitive processes, that is, the indicators of the sets in $\mathcal{F}^{pp}$.

Let F be the indicator of $H \times (a, b]$ with H in $\mathcal{F}_a$. Then, the left side of 6.6 is equal to

$$\mathbb{E}\ 1_H \cdot (N_b - N_a) = \mathbb{E}\ 1_H\ \mathbb{E}_a(N_b - N_a) = \mathbb{E}\ 1_H(\nu_b - \nu_a),$$

where the first equality uses the assumption that $H \in \mathcal{F}_a$ and the second equality uses 6.3. The last member is equal to the right side of 6.6 for the present F. Similarly, 6.6 holds when F is the indicator of $H \times \{0\}$ with H in $\mathcal{F}_0$. □

The following corollary enhances the preceding theorem and provides an example with further uses.

6.7 Corollary. *Let F be a positive predictable process. Let S and T be stopping times with $S \le T$. Then,*

6.8
$$\mathbb{E}_S \int_{(S,T]} F_t\ dN_t = \mathbb{E}_S \int_{(S,T]} F_t\ d\nu_t.$$

Proof. It is enough to show that, for V in $\mathcal{F}_S$ and positive,

$$\mathbb{E}\ V \int_{(S,T]} F_t\ dN_t = \mathbb{E}\ V \int_{(S,T]} F_t\ d\nu_t,$$

which is in turn equivalent to showing that

6.9
$$\mathbb{E} \int_{\mathbb{R}_+} G_t F_t\ dN_t = \mathbb{E} \int_{\mathbb{R}_+} G_t F_t\ d\nu_t,$$

where

$$G_t = V\ 1_{(S,T]}(t), \quad t \in \mathbb{R}_+.$$

The process G is obviously left-continuous, and each G_t is in $\mathcal{F}_t$ by Theorem 1.16d applied to the variables $V 1_{\{S<t\}}$ and $1_{\{t \le T\}}$ separately. It follows from Proposition 6.2 that G is predictable, and thus, so is the product GF. Hence, 6.9 follows from the preceding theorem applied with the positive predictable process GF. □

The preceding theorem and corollary are in fact equivalent to the following theorem about the martingale $\tilde{N}$ defined by 6.4.

6.10 THEOREM. *Let F be a bounded predictable process. Then,*

$$M_t = \int_{[0,t]} F_s \; d\tilde{N}_s, \quad t \in \mathbb{R}_+,$$

is a martingale.

Proof. It is obvious that M is adapted. Each M_t is integrable since $|M_t| \le (N_t + \nu_t)b$ if b is a bound for $|F|$. To show the martingale property that $\mathbb{E}_s(M_t - M_s) = 0$ for $s < t$, it is sufficient to show that

6.11 $$\mathbb{E}_s \int_{(s,t]} F_u \; dN_u = \mathbb{E}_s \int_{(s,t]} F_u \; d\nu_u \; ;$$

this is because $\tilde{N} = N - \nu$ and both sides of 6.11 are real-valued. Now, 6.11 is immediate from Corollary 6.7 applied first with F^+ and then with F^-. □

REMARK. *Stochastic integrals.* The preceding theorem remains true for arbitrary martingales $\tilde{N}$, except that the proof above is no longer valid and, worse, the integral defining M has to be given a new meaning. Above, since $\tilde{N}$ has paths of finite variation over bounded intervals, the integral defining M makes sense ω by ω, that is,

6.12 $$M_t(\omega) = \int_{[0,t]} F_s(\omega) \; d\tilde{N}_s(\omega).$$

But, if $\tilde{N}$ were an arbitrary martingale or, more specifically, if $\tilde{N}$ were a Wiener process, then the paths $s \mapsto \tilde{N}_s(\omega)$ would necessarily have infinite variation over most intervals and, hence, the integral 6.12 has no meaning as a Stieltjes integral for general F. Stochastic calculus goes around the problem by defining M as the limit in probability of a sum of primitive integrals. With this new meaning for the integral M, the conclusion of the last theorem remains true. The interested reader should see a book on stochastic calculus.

Martingale characterization of Poisson processes

Here, we prove Theorem 2.23. The necessity part was already done preceding 2.23. For easy reference, we list next what is to be proved.

6.13 PROPOSITION. *Let N be a counting process adapted to $\mathcal{F}$. Suppose that, for some constant c in $(0, \infty)$,*

$$\tilde{N}_t = N_t - ct, \quad t \in \mathbb{R}_+,$$

is a martingale. Then, N is a Poisson process with rate c.

We start with a lemma of independent interest; it exploits the counting nature of N. Here, $N_{s-} = \lim_{r \uparrow s} N_r$ as usual.

6.14 LEMMA. *Let f be a bounded function on $\mathbb{N}$. Then,*

$$f(N_t) = f(0) + \int_{(0,t]} [f(N_{s-} + 1) - f(N_{s-})] \, dN_s.$$

Proof. Fix an ω. If $N_t(\omega) = 0$ then the claim is obvious. If $N_t(\omega) = n \geq 1$, let $t_1, \ldots, t_n$ be the successive jump times of $s \mapsto N_s(\omega)$ during $(0, t]$, suppressing their dependence on ω. At t_i, the counting function $s \mapsto N_s(\omega)$ jumps from the left-hand limit $i-1$ to the right-hand value i. Thus, the right side of the claimed equation is equal to, for this ω,

$$f(0) + \sum_{i=1}^{n} [f(i - 1 + 1) - f(i - 1)] = f(n) = f(N_t(\omega)).$$

□

Proof of Proposition 6.13. This is an application of Corollary 6.7 with a carefully chosen F. Fix times $s < t$, fix r in $\mathbb{R}_+$, and let

$$f(n) = 1 - e^{-rn}, \quad n \in \mathbb{N}; \quad F_u = f(N_{u-} + 1) - f(N_{u-}), \quad u \in \mathbb{R}_+.$$

Since $u \mapsto N_{u-}$ is adapted and left-continuous, so is F. Thus, F is bounded, positive, and predictable, the last following from Proposition 6.2. Hence, by Corollary 6.7 applied with the current N and $\nu_t = ct$,

$$\mathbb{E}_s \int_{(s,t]} F_u \, dN_u = c \, \mathbb{E}_s \int_{(s,t]} F_u \, du.$$

The integral on the left is equal to $f(N_t) - f(N_s)$ by Lemma 6.14. As to the Lebesgue integral on the right side, replacing F_u with $F_{u+} = f(N_u + 1) - f(N_u)$ will not change the integral since $F_u = F_{u+}$ for all u in $(s, t]$ except for finitely many u. Hence,

$$\mathbb{E}_s[f(N_t) - f(N_s)] = c \, \mathbb{E}_s \int_{(s,t]} [f(N_u + 1) - f(N_u)] \, du.$$

Now we replace t with $s + t$, recall that $f(n) = 1 - e^{-rn}$, and multiply both sides by $\exp rN_s$. The result is (writing $\exp_- x$ for e^{-x})

6.15 $$\mathbb{E}_s \, \exp_- r(N_{s+t} - N_s) = 1 - c(1 - e^{-r})\mathbb{E}_s \int_0^t du \, \exp_- r(N_{s+u} - N_s).$$

Let the left side be denoted by $g(t)$, suppressing its dependence on r, s, ω. We have

$$g(t) = 1 - c(1 - e^{-r}) \int_0^t g(u) \, du,$$

whose only solution is

$$g(t) = \exp_ \, ct(1 - e^{-r}) = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} \; e^{-rk}$$

totally free of s and ω. We have shown that

$$\mathbb{E}_s \exp_ \, r(N_{s+t} - N_s) = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} \; e^{-rk}.$$

Since this is true for arbitrary r in $\mathbb{R}_+$, we conclude that $N_{s+t} - N_s$ is independent of $\mathcal{F}_s$ and has the Poisson distribution with mean ct. This concludes the proof hat N is a Poisson process with rate c. □

6.16 Remark. *Strong Markov property.* The preceding proof can be modified to show that, for the process N,

$$\mathbb{E}_S \; \exp_ \; r \cdot (N_{S+t} - N_S) = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} \; e^{-rk},$$

that is, for every finite stopping time S, the future increment $N_{S+t} - N_S$ is independent of $\mathcal{F}_S$ and has the same Poisson distribution as N_t has. This is called the *strong Markov property* for N. To show it, replace s by S and t by $S + t$ from the beginning of the proof of 6.13, and note that Corollary 6.7 applies full force. This brings us to 6.15 with s replaced by S; and the rest is exactly the same but with s replaced by S.

Non-stationary Poisson processes

These are defined just as Poisson processes except that the distribution of $N_{s+t} - N_s$ has the Poisson distribution with mean $\nu_{s+t} - \nu_s$, where ν is an arbitrary continuous increasing function (the stationary case is where $\nu_t = ct$). In other words, a counting process N adapted to $\mathcal{F}$ is said to be a (*non-stationary*) *Poisson process with mean* ν if ν is continuous increasing real-valued, and, for every positive function f on $\mathbb{N}$,

6.17
$$\mathbb{E}_s f(N_t - N_s) = \sum_{k=0}^{\infty} \frac{e^{-a} a^k}{k!} \; f(k)$$

with $a = \nu_t - \nu_s$; compare this with 6.16 and 2.20. Of course, then, $N - \nu$ is a martingale. The following states this and adds a converse.

6.18 Theorem. *Let N be a counting process adapted to $\mathcal{F}$, and let ν be a (deterministic) increasing continuous real-valued function on $\mathbb{R}_+$ with $\nu_0 = 0$. Then, N is a Poisson process with mean function ν if and only if $N - \nu$ is a martingale.*

Proof. Necessity is trivial. We prove the sufficiency. Given ν, let $\nu_\infty = \lim_{t\to\infty} \nu_t$ and let τ be the functional inverse of ν, that is,

6.19 $$\tau_u = \inf\{t > 0 : \nu_t > u\}, \quad u < \nu_\infty;$$

see Exercise 5.13 of Chapter I. Then, τ is right-continuous and strictly increasing on $[0, \nu_\infty)$, and $\nu_{\tau_u} = u$ by the continuity of ν. Clearly, (N_{τ_u}) is adapted to the filtration $(\mathcal{F}_{\tau_u})$ and is again a counting process. Since $N - \nu$ is assumed to be an $\mathcal{F}$-martingale, the process $(N_{\tau_u} - u)$ is an $(\mathcal{F}_{\tau_u})$-martingale on $[0, \nu_\infty)$. By Proposition 6.13, then, the process N_τ is a Poisson process with rate 1 on the interval $[0, \nu_\infty)$, that is, for every positive function f on $\mathbb{N}$,

6.20 $$\mathbb{E}_{\tau_u} f(N_{\tau_v} - N_{\tau_u}) = \sum_{k=0}^{\infty} \frac{e^{-(v-u)}(v-u)^k}{k!} \, f(k)$$

for $0 \le u < v < \nu_\infty$. There remains to undo the time change 6.19.

We start by observing that, if $\nu_s = \nu_t$ for some $s < t$, then $\mathbb{E}\,(N_t - N_s) = \nu_t - \nu_s = 0$ and thus $N_t - N_s = 0$ almost surely. It follows that, for $0 \le s < t$,

$$N_s = N(\tau(\nu_s)), \quad N_t = N(\tau(\nu_t))$$

almost surely. Hence, taking $u = \nu_s$ and $v = \nu_t$ in 6.20, and putting $a = v - u = \nu_t - \nu_s$, we get

$$\mathbb{E}_{\tau(\nu_s)} f(N_t - N_s) = \sum_{k=0}^{\infty} \frac{e^{-a} a^k}{k!} \, f(k).$$

Finally, apply the conditional expectation operator $\mathbb{E}_s$ on both sides; observing that $\tau(\nu_s) \ge s$ necessarily by the definition 6.19, we get 6.17 with $a = \nu_t - \nu_s$, which completes the proof that N is Poisson with mean ν. □

Martingale characterization of Wiener processes

This is to give the sufficiency part of the proof of Theorem 2.19; recall that the proof of necessity was already given. We list what is to be proved for convenience.

6.21 Proposition. *Let X be a continuous $\mathcal{F}$-martingale with $X_0 = 0$. Suppose that $Y = (X_t^2 - t)_{t\in\mathbb{R}_+}$ is again an $\mathcal{F}$-martingale. Then, X is a Wiener process with respect to $\mathcal{F}$.*

We start by listing a lemma, whose highly technical proof will be given below, after the proof of 6.21.

6.22 Lemma. *Let X be as in 6.21. Let f be a twice differentiable function on $\mathbb{R}$ and suppose that f is bounded along with its derivative f' and its second derivative f''. Then,*

6.23 $$M_t = f \circ X_t - \tfrac{1}{2} \int_0^t ds \; f'' \circ X_s, \quad t \in \mathbb{R}_+,$$

defines a martingale M*.*

Proof of Proposition 6.21. We use the preceding lemma first with $f(x) = \cos rx$ and then with $f(x) = \sin rx$ to conclude that M defined by 6.23 with $f(x) = e^{irx} = \cos rx + i \sin rx$ is a complex-valued martingale. In other words, since $f''(x) = -r^2 f(x)$ when $f(x) = e^{irx}$,

$$\mathbb{E}_s \left[f \circ X_{s+t} - f \circ X_s + \tfrac{1}{2} r^2 \int_s^{s+t} f \circ X_u \, du \right] = 0.$$

Replacing $f(x)$ with e^{irx}, multiplying both sides by $\exp(-irX_s)$, and rearranging the result, we obtain

$$\mathbb{E}_s \exp ir(X_{s+t} - X_s) = 1 - \tfrac{1}{2} r^2 \int_0^t du \; \mathbb{E}_s \exp ir(X_{s+u} - X_s).$$

Let $g(t)$ be defined to be the left side, ignoring its dependence on s and r and ω. We get

$$g(t) = 1 - \tfrac{1}{2} r^2 \int_0^t du \; g(u),$$

whose only solution is $g(t) = \exp(-r^2 t/2)$ independent of s and ω. So, for every r in $\mathbb{R}$,

$$\mathbb{E}_s \exp ir(X_{s+t} - X_s) = \exp(-r^2 t/2),$$

which shows that the increment $X_{s+t} - X_s$ is independent of $\mathcal{F}_s$ and has the Gaussian distribution with mean 0 and variance t; in short, X is Wiener. □

We turn to proving Lemma 6.22. The sophisticated reader will have noticed that it is a simple consequence of Itô's lemma, but we do not have access to such advanced machinery at this point. Instead, the proof is by necessity of the bare-hands type. It is well-worth ignoring it, except for purposes of admiring Lévy and Doob for their power and ingenuity.

Proof of Lemma 6.22. a) We shall eventually show that $\mathbb{E}_a(M_b - M_a) = 0$ for $0 \le a < b < \infty$. Fix a and b such, fix $\varepsilon > 0$, take an integer $n \ge 1$, and let $\delta = (b - a)/n$. Define

$$T = b \wedge \inf\{t \ge a : \max_{a \le p,q \le t, \; |q-p| \le \delta} |X_q - X_p| = \varepsilon\}.$$

Since $t \mapsto X_t(\omega)$ is continuous, it is uniformly continuous on $[a, b]$, and hence, $T(\omega)$ is equal to b for all n large enough, depending on ε and the outcome ω. By the continuity of X,

$$\{T > t\} = \bigcup_m \bigcap_{p,q} \left\{ |X_q - X_p| < \varepsilon - \frac{1}{m} \right\},$$

where the union is over all the integers $m \ge 1$ and the intersection is over all rationals p and q in $[a, t]$ with $|q - p| \le \delta$. Since X is adapted, this shows

that T is a stopping time. Consequently, since X and Y are martingales by hypothesis, so are the processes obtained from them by stopping at T, denoted by

6.24 $$Z = (Z_t) = (X_{t\wedge T}), \quad \bar{Z} = (\bar{Z}_t) = (Z_t^2 - t\wedge T).$$

It follows that, for s and t in $[a, b]$ with $0 < t - s \le \delta$,

6.25 $$\mathbb{E}_s(Z_t - Z_s) = 0, \quad |Z_t - Z_s| \le \varepsilon,$$

6.26 $$\mathbb{E}_s(Z_t - Z_s)^2 = \mathbb{E}_s[Z_t^2 - 2Z_s(Z_t - Z_s) - Z_s^2] = \mathbb{E}_s(Z_t^2 - 0 - Z_s^2) = \mathbb{E}_s(t\wedge T - s\wedge T).$$

b) Lef f be as described. We shall use Taylor's theorem in the following form

6.27 $$f(y) - f(x) = f'(x)(y-x) + \frac{1}{2}f''(x)(y-x)^2 + r(x,y),$$

where the remainder term is such that, for some continuous increasing function h on $\mathbb{R}_+$ with $h(0) = 0$,

6.28 $$|r(x,y)| \le (y-x)^2 h(|y-x|).$$

c) Keeping $a, b, \varepsilon, n, \delta$ as before, let $\mathcal{D}$ be the subdivision of the interval $(a, b]$ into n disjoint intervals of equal length and of form $(s, t]$, that is,

$$\mathcal{D} = \{(s,t] : \quad s = a + k\delta, t = s + \delta, k = 0, 1, \ldots, n-1\}.$$

Using 6.27, with $\sum_{\mathcal{D}}$ indicating summation over all $(s, t]$ in $\mathcal{D}$,

6.29 $$f\circ Z_b - f\circ Z_a = \sum_{\mathcal{D}} [f\circ Z_t - f\circ Z_s] = \sum_{\mathcal{D}} [(f'\circ Z_s)(Z_t - Z_s) + \frac{1}{2}(f''\circ Z_s)(Z_t - Z_s)^2] + R,$$

where the remainder term R satisfies, in view of 6.25 and 6.28,

6.30 $$|R| \le \sum_{\mathcal{D}} (Z_t - Z_s)^2 \; h\circ|Z_t - Z_s| \le h(\varepsilon) \sum_{\mathcal{D}} (Z_t - Z_s)^2.$$

We now apply $\mathbb{E}_a$ to both sides of 6.29 and 6.30, using $\mathbb{E}_a\mathbb{E}_s = \mathbb{E}_a$ repeatedly for $a \le s$ and using the equalities in 6.25 and 6.26. We get

6.31 $$\mathbb{E}_a f\circ Z_b - f\circ Z_a = \tfrac{1}{2}\mathbb{E}_a \textstyle\sum_{\mathcal{D}} (f''\circ Z_s)(t\wedge T - s\wedge T) + \mathbb{E}_a R,$$

6.32 $$|\mathbb{E}_a R| \le h(\varepsilon)\mathbb{E}_a \textstyle\sum_{\mathcal{D}} (t\wedge T - s\wedge T) \le h(\varepsilon)\mathbb{E}_a(b-a) \le h(\varepsilon)(b-a).$$

Consider the sum over $(s, t]$ in $\mathcal{D}$ on the right side of 6.31. For $(q, r]$ in $\mathcal{D}$, on the event $\{q < T \le r\}$, we have $Z_s = X_s$ for $s \le q$ and the sum is equal to

$$\delta \sum_{\mathcal{D}} f''\circ X_s - Q$$

where

$$|Q| = |(f'' \circ X_q)(r-T) + \delta \sum_{s>q,\mathcal{D}} f'' \circ X_s| \le \|f''\| \cdot (b-T)$$

with $\|f''\| = \sup_x |f''(x)| < \infty$ by assumption. Hence, recalling that $a \le T \le b$ we can re-write 6.31 and 6.32 as follows:

6.33 $$\mathbb{E}_a f \circ X_{T\wedge b} - f \circ X_a = \tfrac{1}{2}\mathbb{E}_a \textstyle\sum_{\mathcal{D}} \delta f'' \circ X_s - \tfrac{1}{2}\mathbb{E}_a Q + \mathbb{E}_a R,$$

6.34 $$|\mathbb{E}_a Q| \le \|f''\| \mathbb{E}_a (b-T), \quad |\mathbb{E}_a R| \le (b-a)h(\varepsilon).$$

d) Keeping a, b, ε as before, we now let $n \to \infty$. Then, T increases to b as mentioned earlier. So, $T \wedge b \to b$, and $X_{T\wedge b} \to X_b$ by the continuity of X, which implies that $f \circ X_{T\wedge b} \to f \circ X_b$ by the continuity of f, and hence,

6.35 $$\mathbb{E}_a f \circ X_{T\wedge b} \to \mathbb{E}_a f \circ X_b$$

by the bounded convergence theorem (recall that f is bounded). Again as $n \to \infty$, on the right side of 6.33, the sum over $\mathcal{D}$ converges to the Riemann integral of $f'' \circ X_s$ over $[a, b]$, and

6.36 $$\mathbb{E}_a \sum_{\mathcal{D}} \delta f'' \circ X_s \to \mathbb{E}_a \int_a^b f'' \circ X_u \, du$$

by the bounded convergence theorem, since the sum remains dominated by $\|f''\| \cdot (b-a)$ for all n. Finally, since T increases to b, $\mathbb{E}_a(b-T) \to 0$ by the bounded convergence theorem, which yields

6.37 $$|\mathbb{E}_a Q| \to 0$$

in view of 6.34. Putting 6.35, 6.36, 6.37 into 6.33 and noting the bound for $\mathbb{E}_a R$ in 6.34, we obtain

$$|\mathbb{E}_a(M_b - M_a)| = \left| \mathbb{E}_a f \circ X_b - f \circ X_a - \frac{1}{2}\mathbb{E}_a \int_a^b du \; f'' \circ X_u \right| \le (b-a)h(\varepsilon).$$

This shows that M is a martingale, since a, b, ε are arbitrary and $h(\varepsilon)$ decreases to 0 as $\varepsilon \to 0$. □

7 Standard Filtrations and Modifications of Martingales

This section is to supplement the chapter by discussing the right-continuity and augmentation of filtrations, and the beneficial consequences of such properties on stopping times and martingales. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is the probability space in the background, the time-set is $\mathbb{R}_+$ unless specified otherwise, and all filtrations and processes are indexed by $\mathbb{R}_+$.

Augmentation

Let $\mathcal{F} = (\mathcal{F}_t)$ be a filtration. It is said to be *augmented* if $(\Omega, \mathcal{H}, \mathbb{P})$ is complete and all the negligible events in $\mathcal{H}$ are also in $\mathcal{F}_0$ (and, therefore, in every $\mathcal{F}_t$).

Suppose that $(\Omega, \mathcal{H}, \mathbb{P})$ is complete. Let $\mathcal{F}$ be an arbitrary filtration. Let $\mathcal{N}$ be the collection of all negligible events in $\mathcal{H}$ and let $\bar{\mathcal{F}}_t$ be the σ-algebra generated by $\mathcal{F}_t \cup \mathcal{N}$. Then, $\bar{\mathcal{F}} = (\bar{\mathcal{F}}_t)$ is an augmented filtration and is called the *augmentation* of $\mathcal{F}$. Obviously, $\mathcal{F}$ is augmented if and only if $\mathcal{F} = \bar{\mathcal{F}}$.

Right-continuity

Let $\mathcal{F}$ be a filtration. We define

7.1 $$\mathcal{F}_{t+} = \cap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}, \quad t \in \mathbb{R}_+.$$

Then, $(\mathcal{F}_{t+})$ is again a filtration and is finer than $(\mathcal{F}_t)$. The filtration $\mathcal{F}$ is said to be right-continuous if

7.2 $$\mathcal{F}_t = \mathcal{F}_{t+}$$

for every t in $\mathbb{R}_+$. Note that $(\mathcal{F}_{t+})$ is itself a right-continuous filtration; it is the coarsest right-continuous filtration that is finer than $\mathcal{F}$.

Heuristically, $\mathcal{F}_{t+}$ has the same information as $\mathcal{F}_t$ plus the information gained by an "infinitesimal peek" into the future. For instance if $t \mapsto X_t$ depicts a smooth motion and $\mathcal{F}$ is the filtration generated by $X = (X_t)$, then $\mathcal{F}_t$ has all the information regarding the past of X and the present position X_t, whereas $\mathcal{F}_{t+}$ has all that information plus the velocity $V_t = \lim_{\varepsilon\to 0}(X_{t+\varepsilon} - X_t)/\varepsilon$ and acceleration at t and so on.

Sometimes, the difference between $(\mathcal{F}_{t+})$ and $(\mathcal{F}_t)$ is so slight that the augmentation of $(\mathcal{F}_t)$ is right-continuous and therefore finer than $(\mathcal{F}_{t+})$. We shall see several instances of it, especially with Lévy processes and Brownian motion. The following illustrates this in a simple case.

7.3 Example. Let T and V be positive random variables with diffuse distributions on $\mathbb{R}_+$. Define

$$X_t(\omega) = \begin{cases} V(\omega)t & \text{if } t < T(\omega), \\ V(\omega)T(\omega) & \text{if } t \geq T(\omega). \end{cases}$$

The process $X = (X_t)$ describes the motion of a particle that starts at the origin at time 0, moves with speed V until the time T, and stops at time T. Let $\mathcal{F}$ be the filtration generated by X. Note that T is not a stopping time of $\mathcal{F}$, the reason being that knowing $X_s(\omega) = V(\omega)s$ for all $s \leq t$ is insufficient to tell whether $T(\omega) = t$ or $T(\omega) > t$. But, the event $\{T \leq t\}$ is definitely in $\mathcal{F}_{t+\varepsilon}$ for every $\varepsilon > 0$, and thus, is in $\mathcal{F}_{t+}$; in other words, T is a stopping time of $(\mathcal{F}_{t+})$. The failure of T to be a stopping time of $(\mathcal{F}_t)$ is due to a negligible cause: the event $\{T = t\}$ is negligible by our assumption that the distribution of T is diffuse. Hence, letting $(\bar{\mathcal{F}}_t)$ be the augmentation of $(\mathcal{F}_t)$, we conclude that T is a stopping time of $(\bar{\mathcal{F}}_t)$ and that $\bar{\mathcal{F}}_t \supset \mathcal{F}_{t+}$ for all t except $t = 0$.

Stopping times and augmentation

Let $(\mathcal{F}_t)$ be a filtration and let T be a random time, a mapping from Ω into $\bar{\mathbb{R}}_+$. Suppose that $\mathcal{F}$ is augmented and S is a stopping time of it. If $T = S$ almost surely, then T is a stopping time of $\mathcal{F}$: For each time t, the event $\{S \le t\}$ belongs to $\mathcal{F}_t$, and $\{T \le t\}$ and $\{S \le t\}$ differ from each other by negligible events, and those negligible events are in $\mathcal{F}_t$ by our assumption that $\mathcal{F}$ is augmented.

Stopping times and right-continuity

Right-continuity of a filtration simplifies the tests for its stopping times. This is a corollary of the following.

7.4 Theorem. *Let $\mathcal{F}$ be a filtration, and T a random time. Then, T is a stopping time of $(\mathcal{F}_{t+})$ if and only if*

7.5 $$\{T < t\} \in \mathcal{F}_t \text{ for every } t \text{ in } \mathbb{R}_+.$$

Proof. Let $\varepsilon_n = 1/n$, $n \ge 1$. If 7.5 holds, then

$$\{T \le t\} = \cap_n \{T < t + \varepsilon_n\} \in \cap_n \mathcal{F}_{t+\varepsilon_n} = \mathcal{F}_{t+}$$

for every t, which means that T is a stopping time of $(\mathcal{F}_{t+})$. Conversely, if T is a stopping times of $(\mathcal{F}_{t+})$, then $\{T \le s\} \in \mathcal{F}_{s+} \subset \mathcal{F}_t$ for all $s < t$, and hence,

$$\{T < t\} = \cup_n \{T \le t - \varepsilon_n\} \in \mathcal{F}_t. \qquad \square$$

If $\mathcal{F}$ is right-continuous, then $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, and the preceding theorem shows that T is a stopping time of $\mathcal{F}$ if and only if 7.5 holds.

7.6 Example. Let W be a Wiener process and let $\mathcal{F}$ be the filtration generated by it. For fixed $a > 0$, let

$$T = \inf\{t > 0 : \ W_t > a\}.$$

Then, T is not a stopping time of $(\mathcal{F}_t)$, but it is a stopping time of $(\mathcal{F}_{t+})$. The latter assertion follows from the preceding theorem, because $T(\omega) < t$ if and only if $W_r(\omega) > a$ for some positive rational number $r < t$. The former assertion follows from observing that $T(\omega) = t$ if and only if $W_r(\omega) \le a$ for all rationals $r < t$ and $r = t$ and $W_r(\omega) > a$ for some rational $r > t$, and the last part with $r > t$ cannot be told at the time t.

Past until T

Let $\mathcal{F}$ be a filtration. Let T be a stopping time of the filtration $(\mathcal{F}_{t+})$, and let $\mathcal{F}_{T+}$ denote the corresponding past until T, that is, $\mathcal{F}_{T+} = \mathcal{G}_T$ where

$(\mathcal{G}_t) = (\mathcal{F}_{t+})$. More explicitly , recalling the definition 1.9 with the filtration $(\mathcal{G}_t)$,

7.7 $$\mathcal{F}_{T+} = \{H \in \mathcal{H} : \ H \cap \{T \le t\} \in \mathcal{F}_{t+} \text{ for every } t \text{ in } \mathbb{R}_+\}.$$

In fact, the arguments of the proof of 7.4 shows that

7.8 $$\mathcal{F}_{T+} = \{H \in \mathcal{H} : \ H \cap \{T < t\} \in \mathcal{F}_{t} \text{ for every } t \text{ in } \mathbb{R}_+\}.$$

Of course, if $\mathcal{F}$ is right-continuous, 7.7 shows that $\mathcal{F}_{T+} = \mathcal{F}_T$, and 7.8 becomes another characterization for $\mathcal{F}_T$.

Sequences of stopping times

Let (T_n) be a sequence of stopping times of a filtration $\mathcal{F}$. If the sequence is increasing, its limit is a stopping time of $\mathcal{F}$; see Exercise 1.34. The following contains the best that can be said about decreasing sequences.

7.9 PROPOSITION. *Let (T_n) be a sequence of stopping times of $(\mathcal{F}_t)$ or of $(\mathcal{F}_{t+})$. Then, $T = \inf T_n$ is a stopping time of $(\mathcal{F}_{t+})$, and*

$$\mathcal{F}_{T+} = \cap_n \mathcal{F}_{T_n+}.$$

Proof. Since $(\mathcal{F}_t)$ is coarser than $(\mathcal{F}_{t+})$, every stopping time of the former is a stopping time of the latter. So, the T_n are stopping times of $(\mathcal{F}_{t+})$ in either case. By Theorem 7.4, the event $\{T_n < t\}$ is in $\mathcal{F}_t$ for every n and every t. It follows that

$$\{T < t\} = \cup_n \{T_n < t\} \in \mathcal{F}_t$$

for every t, that is, T is a stopping time of $(\mathcal{F}_{t+})$ in view of 7.4.

Since $T \le T_n$ for every n, Theorem 1.16b applied with the filtration $(\mathcal{F}_{t+})$ shows that $\mathcal{F}_{T+} \subset \mathcal{F}_{T_n+}$ for every n. Hence, $\mathcal{F}_{T+} \subset \cap_n \mathcal{F}_{T_n+}$. To show the converse, let H be an event that belongs to $\mathcal{F}_{T_n+}$ for every n. Then,

$$H \cap \{T < t\} = \cup_n (H \cap \{T_n < t\}) \in \mathcal{F}_t$$

in view of 7.8 applied with T_n. Thus, by 7.8 again, $H \in \mathcal{F}_{T+}$. □

If $\mathcal{F}$ is right-continuous, and (T_n) is a sequence of stopping times of it, then the infimum and supremum and limit inferior and limit superior of the sequence are all stopping times.

Times foretold

Let $\mathcal{F}$ be a filtration. Recall that a random time T is said to be *foretold* by a stopping time S of $\mathcal{F}$ if $T \ge S$ and $T \in \mathcal{F}_S$. Obviously, if S is a stopping time of $(\mathcal{F}_{t+})$ and $T \ge S$ and $T \in \mathcal{F}_{S+}$, then T is a stopping time of $(\mathcal{F}_{t+})$. The following is a refinement.

7.10 Proposition. *Let S be a stopping time of $(\mathcal{F}_{t+})$. Let T be a random time such that $T \in \mathcal{F}_{S+}$ and $T \geq S$, with strict inequality $T > S$ on the event $\{S < \infty\}$. Then, T is a stopping time of $(\mathcal{F}_t)$ and $\mathcal{F}_{S+} \subset \mathcal{F}_T$.*

Proof. Let S and T be as described. For every outcome ω and time t, if $T(\omega) \leq t$ then $S(\omega) < T(\omega)$ and $S(\omega) < t$. Thus, for H in $\mathcal{H}$,

7.11 $$H \cap \{T \leq t\} = H \cap \{T \leq t\} \cap \{S < t\}.$$

Suppose that $H \in \mathcal{F}_{S+}$. Since $T \in \mathcal{F}_{S+}$ by assumption, then, the left side is in $\mathcal{F}_{S+}$, which implies that the right side is in $\mathcal{F}_t$ in view of 7.8 for $\mathcal{F}_{S+}$. Thus, the left side of 7.11 is in $\mathcal{F}_t$ whenever $H \in \mathcal{F}_{S+}$. Taking $H = \Omega$ shows that T is a stopping time of $(\mathcal{F}_t)$, and we conclude that every H in $\mathcal{F}_{S+}$ is in $\mathcal{F}_T$. □

Approximation of stopping times

The following shows that Proposition 1.20 remains true for stopping times T of $(\mathcal{F}_{t+})$. Here, the d_n are as defined by 1.19.

7.12 Proposition. *Let T be a stopping time of $(\mathcal{F}_{t+})$. Let $T_n = d_n \circ T$ for each n in $\mathbb{N}$. Then, each T_n is a stopping time of $(\mathcal{F}_t)$, each T_n is discrete, and the sequence (T_n) decreases to T and decreases strictly on the set $\{T < \infty\}$. Moreover, $\mathcal{F}_{T+} = \cap_n \mathcal{F}_{T_n}$.*

Proof. The proof is immediate from Propositions 7.9 and 7.10 once we note that each T_n is foretold by T. □

Hitting times

Augmented right-continuous filtrations are desirable for the simplifications noted above and for the following important result, which we list here without proof.

Let $X = (X_t)$ be a stochastic process with state space $(E, \mathcal{E})$, where E is topological and $\mathcal{E}$ is the Borel σ-algebra on E. Let $\mathcal{F}$ be a filtration.

7.13 Theorem. *Suppose that $\mathcal{F}$ is right-continuous and augmented. Suppose that X is right-continuous and is adapted to $\mathcal{F}$. Then, for every Borel subset B of E, the hitting time*

$$T_B = \inf\{t \in \mathbb{R}_+ : \ X_t \in B\}$$

is a stopping time of $\mathcal{F}$.

Regularity of martingales

We start with a filtration $\mathcal{F}$ on $\mathbb{R}_+$. In the following, D is an arbitrary countable subset of $\mathbb{R}_+$ which is dense in $\mathbb{R}_+$, for example, one can take D to be the set of all rationals in $\mathbb{R}_+$.

7.14 PROPOSITION. *Let X be an $\mathcal{F}$-submartingale on $\mathbb{R}_+$. For almost every ω, the following limits exist and are in $\mathbb{R}$:*

7.15 $$X_{t+}(\omega) = \lim_{r\in D,\ r\downarrow t} X_r(\omega), \quad t \geq 0,$$

7.16 $$X_{t-}(\omega) = \lim_{r\in D,\ r\uparrow t} X_r(\omega), \quad t > 0.$$

Proof. Fix s in D. Let a and b be rational numbers with $a < b$. Let B be a finite subset of $D \cap [0, s]$ that includes the end point s. Then, $(X_r)_{r\in B}$ is a submartingale with respect to $(\mathcal{F}_r)_{r\in B}$ with a discrete-time set B. By Theorem 3.21 applied to the submartingale X on B,

7.17 $$c\ \mathbb{P}\{\max_{r\in B} X_r \geq c\} \leq \mathbb{E}\ |X_s|.$$

Next, let $U_B(a, b)$ be the number of upcrossings of the interval (a, b) by the process $(X_r)_{r\in B}$. By Proposition 3.19,

7.18 $$(b - a)\mathbb{E}\ U_B(a, b) \leq \mathbb{E}\ (X_s - a)^+ < \infty.$$

Note that the right sides of 7.17 and 7.18 are free of B. Thus, by taking supremums over all finite sets B that include s and are contained in $D\cap[0, s]$, we see that the same inequalities hold for

$$M_s = \sup_{r\in D\cap[0,s]} |X_r|, \quad U_s(a, b) = \sup_B U_B(a, b)$$

respectively. It follows that $M_s < \infty$ and $U_s(a, b) < \infty$ almost surely.

For s in D, let Ω_s be the collection of all ω for which the limits $X_{t-}(\omega)$ exist and are finite for all t in $(0, s]$ and the limits $X_{t+}(\omega)$ exist and are finite for all t in $[0, s)$. Observe that

$$\Omega_s \supset \cap_{a,b}\{M_s < \infty, U_s(a, b) < \infty\},$$

where the intersection is over all pairs (a, b) of rationals with $a < b$. For each s in D, this shows that Ω_s contains an almost sure event. Hence, $\Omega_0 = \cap_{s\in D}\Omega_s$ contains an almost sure event, and for every ω in Ω_0 the limits indicated exist and are in $\mathbb{R}$. □

7.19 PROPOSITION. *Suppose that $\mathcal{F}$ is right-continuous and augmented, and let X be a $\mathcal{F}$-submartingale. Let Ω_0 be the almost sure set of all ω for which the limits 7.15 and 7.16 exist in $\mathbb{R}$, and set $X_{t-}(\omega) = X_{t+}(\omega) = 0$ for every ω outside Ω_0.*

a) For each t in $\mathbb{R}_+$, the random variable X_{t+} is integrable and

$$X_t \leq X_{t+}$$

almost surely. Here, the equality holds almost surely if and only if $s \mapsto \mathbb{E}\ X_s$ is right-continuous at t (in particular, if X is a martingale).

b) The process $(X_{t+})_{r\in\mathbb{R}_+}$ is a submartingale with respect to $\mathcal{F}$, and it is a martingale if X is so. Moreover, for every outcome ω, the trajectory $t \mapsto X_{t+}(\omega)$ is right-continuous and left-limited.

Proof. Since $\mathcal{F}$ is augmented, the set Ω_0 belongs to $\mathcal{F}_0$ and to $\mathcal{F}_t$ for all t; thus the alteration outside Ω_0 does not change the adaptedness of X to $\mathcal{F}$, and the altered X is still an $\mathcal{F}$-submartingale.

a) Fix t in $\mathbb{R}_+$. Let (r_n) be a sequence in D decreasing strictly to t. Then, (X_{r_n}) is a reversed time submartingale, and $\mathbb{E}\, X_t \le \mathbb{E}\, X_{r_n}$ for every n. By Theorem 4.19, the sequence (X_{r_n}) is uniformly integrable and converges to X_{t+} almost surely and in L^1. It follows that X_{t+} is integrable and, for every event H in $\mathcal{F}_t$,

7.20 $$\mathbb{E}\, X_{t+}1_H = \lim\ \mathbb{E}\, X_{r_n}1_H \ge \mathbb{E}\, X_t 1_H,$$

the inequality being through the submartingale inequality for $t < r_n$. Thus,

7.21 $$\mathbb{E}_t(X_{t+} - X_t) \ge 0.$$

Since X_{r_n} belongs to $\mathcal{F}_{t+\varepsilon}$ for every $\varepsilon > 0$ and all n large enough, the limit X_{t+} belongs to $\mathcal{F}_{t+}$, and $\mathcal{F}_{t+} = \mathcal{F}_t$ by the assumed right-continuity for $\mathcal{F}$. Thus, the left side of 7.21 is equal to $X_{t+} - X_t$, which proves the claim that $X_{t+} \ge X_t$ almost surely. The equality would hold almost surely if and only if $\mathbb{E}\, X_{t+} = \mathbb{E}\, X_t$, which in turn holds if and only if $s \mapsto \mathbb{E}\, X_s$ is right-continuous at t (in which case $\mathbb{E}\, X_t = \lim\ \mathbb{E}\, X_{r_n} = \mathbb{E}\, X_{t+}$).

b) The paths $t \mapsto X_{t+}(\omega)$ are right-continuous and left-limited by the way they are defined. To see that (X_{t+}) is an $\mathcal{F}$-submartingale, take $s < t$, choose (r_n) in D strictly decreasing to t, and (q_n) in D strictly decreasing to s, ensuring that $s < q_n < t < r_n$ for every n. Then, for H in $\mathcal{F}_s$, using 7.20 twice, we get

$$\mathbb{E}\, X_{s+}1_H = \lim\ \mathbb{E}\, X_{q_n}1_H \le \lim\ \mathbb{E}\, X_{r_n}1_H = \mathbb{E}\, X_{t+}1_H,$$

where the inequality is through the submartingale property of X. This completes the proof. □

Modifications of martingales

The following is an immediate corollary of the last theorem: put $Y_t(\omega) = X_{t+}(\omega)$ for every t and every ω.

7.22 THEOREM. *Suppose that $\mathcal{F}$ is right-continuous and augmented. Suppose that X is an $\mathcal{F}$-submartingale and $t \mapsto \mathbb{E}\, X_t$ is right-continuous. Then, there exists a process Y such that*

a) for every ω, the path $t \mapsto Y_t(\omega)$ is right-continuous and left-limited,

b) *Y is an $\mathcal{F}$-submartingale,*
c) *for every t in $\mathbb{R}_+$, we have $\mathbb{P}\{X_t = Y_t\} = 1$.*

The preceding theorem is the justification for limiting our treatment in Section 5 to right-continuous processes. Note that, if X is a martingale, the right-continuity of $\mathbb{E}\ X_t$ in t is immediate, since $\mathbb{E}\ X_t = \mathbb{E}\ X_0$ for all t. The process Y is said to be a *modification* of X because of the statement (c), which matter we clarify next.

Modifications and indistinguishability

Let $\mathbb{T}$ be some index set. Let $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$ be stochastic processes with the same state space. Then, one is said to be a *modification* of the other if, for each t in $\mathbb{T}$,

$$\mathbb{P}\{X_t = Y_t\} = 1.$$

They are said to be *indistinguishable* if

$$\mathbb{P}\{X_t = Y_t \text{ for all } t \text{ in } \mathbb{T}\} = 1.$$

For example, in Theorem 7.22, the last assertion is that Y is a modification of X; it does not mean that they are indistinguishable.

Suppose that X is a modification of Y. Then, for every integer $n < \infty$ and indices $t_1, \ldots, t_n$ in $\mathbb{T}$, the vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(Y_{t_1}, \ldots, Y_{t_n})$ are almost surely equal. It follows that the two vectors have the same distribution. In other words, X and Y have the same finite-dimensional distributions, and therefore, they have the same probability law.

If X and Y are indistinguishable, then they are modifications of each other. If they are modifications of each other, and if the index set $\mathbb{T}$ is countable, then they are indistinguishable. Otherwise, in general, indistinguishability requires more than being modifications of each other.

For instance, suppose $\mathbb{T} = \mathbb{R}_+$ and the state space is $\mathbb{R}^d$. Suppose that X and Y are modifications of each other and are both right-continuous. Then they are indistinguishable.

Exercises

7.23 *Right-continuity.* In Example 7.3, describe $\mathcal{F}_0$ and $\mathcal{F}_{0+}$. Let $\bar{\mathcal{F}}$ be the augmentation of $\mathcal{F}$. Show that $\bar{\mathcal{F}}_{t+} = \bar{\mathcal{F}}_t$ for $t > 0$.

7.24 *Stopping times.* Show that a random time T is a stopping time of $(\mathcal{F}_{t+})$ if and only if the process $(T \wedge t)_{t\in\mathbb{R}_+}$ is adapted to $(\mathcal{F}_t)$.

7.25 *Strict past at T.* Recall from Exercise 1.31 that, for a random time T, the strict past at T is defined to be the σ-algebra generated by events of the form $H \cap \{t < T\}$ with H in $\mathcal{F}_t$ and $t \in \mathbb{R}_+$. Then, T belongs to $\mathcal{F}_{T-}$ and $\mathcal{F}_{T-} \subset \mathcal{F}_T$.

7.26 *Continuation.* Suppose that $\mathcal{F}$ is right-continuous and augmented. Let S and T be stopping times of it. Show that, for every H in $\mathcal{F}_S$, the event $H \cap \{S < T\}$ belongs to $\mathcal{F}_{T-}$. In particular, $\{S < T\} \in \mathcal{F}_{T-}$. Show that $\{S < T\}$ belongs to $\mathcal{F}_S \cap \mathcal{F}_{T-}$.

7.27 *Debuts and stopping times.* Let $\mathcal{F}$ be right-continuous and augmented as in Theorem 7.13. For $A \subset \mathbb{R}_+ \times \Omega$, let

$$D_A(\omega) = \inf\{t \in \mathbb{R}_+ : \ (t, \omega) \in A\}, \quad \omega \in \Omega.$$

If the process (X_t) defined by $X_t(\omega) = 1_A(t, \omega)$ is progressive in the sense of 1.15, then D_A is a stopping time of $\mathcal{F}$. Theorem 7.13 is a special case of this remark.

7.28 *Hitting times.* For fixed $a > 0$, let T be defined as in Example 7.6. Show that $T = T_a$ almost surely, where T_a is as defined by 5.19.

7.29 *Continuation.* Let T_a be defined by 5.19 and let S_a be the T defined in 7.6. Show that $a \mapsto T_a$ is left-continuous and $a \mapsto S_a$ is right-continuous. In fact, $T_a = S_{a-}$, the left-limit at a of S; Show this. The process (S_a) is a right-continuous modification of (T_a). They are eminently distinguishable.

7.30 *Predictable stopping times.* Let $\mathcal{F}$ be right-continuous and augmented. Let T be a stopping time of it. Then, T is said to be predictable if the set $\{(\omega, t) : \ t \geq T(\omega)\}$ belongs to the predictable σ-algebra; see 6.1 for the latter. Equivalently, T is said to be *predictable* if there exists a sequence (T_n) of stopping times such that, for every ω for which $T(\omega) > 0$, the sequence of numbers $T_n(\omega)$ increases strictly to $T(\omega)$. If W is a Wiener process and $\mathcal{F}$ is generated by it, then every stopping time of $\mathcal{F}$ is predictable.

7.31 *Classification of stopping times.* In addition, a stopping time T is said to be *σ-predictable* if there exists a sequence of predictable stopping times T_n such that, for every ω, we have $T(\omega) = T_n(\omega)$ for some n. Finally, T is said to be *totally unpredictable* if $\mathbb{P}\{T = S\} = 0$ for every predictable stopping time S. In Example 1.7, suppose that N is a Poisson process. Then, $T_1, T_2, \ldots$ are all totally unpredictable, the time T is predictable. The stopping time $T \wedge T_1$ is neither predictable nor totally unpredictable; it is equal to the predictable time T on the event $\{T < T_1\}$ and to the totally unpredictable time T_1 on the event $\{T > T_1\}$. This example is instructive. In general, for every stopping time T there exist a σ-predictable stopping time R and a totally unpredictable stopping time S such that $T = R \wedge S$. In the case of standard Markov processes, S is a jump time and R is a time of continuity for the process.

Chapter VI

Poisson Random Measures

The aim is to give a reasonably detailed account of Poisson random measures and their uses. Such random measures are often the primitive elements from which more evolved processes are constructed. We shall illustrate several such constructions which yield Lévy processes and pure-jump Markov processes.

Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is the probability space in the background. We shall have many occasions to use the notation $\exp_{-} x$ for e^{-x}.

1 Random Measures

Let $(E, \mathcal{E})$ be a measurable space. A *random measure* on $(E, \mathcal{E})$ is a transition kernel from $(\Omega, \mathcal{H})$ into $(E, \mathcal{E})$.

More explicitly, a mapping $M : \Omega \times \mathcal{E} \mapsto \bar{\mathbb{R}}_+$ is called a random measure if $\omega \mapsto M(\omega, A)$ is a random variable for each A in $\mathcal{E}$ and if $A \mapsto M(\omega, A)$ is a measure on $(E, \mathcal{E})$ for each ω in Ω. We shall denote by $M(A)$ the former random variable; then, we may regard M as the collection of random variables $M(A)$, $A \in \mathcal{E}$. We shall denote by M_ω the latter measure $A \mapsto M(\omega, A)$; the term "random measure" signifies that M is a random variable that assigns a measure M_ω to every outcome ω in Ω.

The terms such as finite, σ-finite, diffuse, etc. are used for a random measure M if the measure M_ω has the corresponding properties for almost every ω in Ω. In particular, it is said to be integer-valued if it takes values in $\bar{\mathbb{N}} = \{0, 1, \ldots, +\infty\}$. It is said to be a *random counting measure* if, for almost every ω, the measure M_ω is purely atomic and its every atom has weight one.

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_6,

Mean measures and integrals

Let M be a random measure on $(E, \mathcal{E})$. Recall that $\mathcal{E}_+$ denotes the set of all positive $\mathcal{E}$-measurable functions. We recall also the results and notations of the measure-kernel-function setup: According to Fubini's theorem, I.6.3,

1.1 $$Mf(\omega) = \int_E M(\omega, dx)\ f(x), \quad \omega \in \Omega,$$

defines a positive random variable Mf for each f in $\mathcal{E}_+$, and

1.2 $$\mu(A) = \mathbb{E}\ M(A) = \int_\Omega \mathbb{P}(d\omega)\ M(\omega, A), \quad A \in \mathcal{E},$$

defines a measure μ on $(E, \mathcal{E})$, called the mean of M, and

1.3 $$\mathbb{E}\ Mf = \mu f, \quad f \in \mathcal{E}_+.$$

Laplace functionals

Let M be a random measure on $(E, \mathcal{E})$. It may be regarded as a collection $\{M(A) : A \in \mathcal{E}\}$ indexed by $\mathcal{E}$ or as a collection $\{Mf : f \in \mathcal{E}_+\}$ indexed by $\mathcal{E}_+$. Thus, to specify the probability law of M, it is sufficient to specify the joint distribution of $Mf_1, \ldots, Mf_n$ for every choice of integer $n \geq 1$ and functions $f_1, \ldots, f_n$ in $\mathcal{E}_+$. Such a joint distribution can be specified implicitly via the joint Laplace transform (recall the notation $\exp_-\, x = e^{-x}$)

$$\mathbb{E}\ \exp_-(r_1 Mf_1 + \cdots + r_n Mf_n), \quad r_1, \ldots, r_n \in \mathbb{R}_+.$$

Noting that this is the same as $\mathbb{E}\ \exp_-\, Mf$ with $f = r_1 f_1 + \cdots + r_n f_n$, we obtain the proof of the following.

1.4 PROPOSITION. *The probability law of a random measure M on $(E, \mathcal{E})$ is determined uniquely by*

$$\mathbb{E}\ e^{-Mf}, \quad f \in \mathcal{E}_+. \qquad \square$$

The mapping $f \mapsto \mathbb{E}\ e^{-Mf}$ from $\mathcal{E}_+$ into $[0,1]$ is called the *Laplace functional* of M. It is continuous under increasing limits:

1.5 PROPOSITION. *If $(f_n) \subset \mathcal{E}_+$ is increasing to f, then*

$$\lim_n \mathbb{E}\ \exp_-\, Mf_n = \mathbb{E}\ \exp_-\, Mf.$$

Proof. If (f_n) is increasing to f, then (Mf_n) is increasing to Mf by the monotone convergence theorem for the measure M_ω applied for each outcome ω, and thus, $(\exp_-\, Mf_n)$ is decreasing to $\exp_-\, Mf$. The desired conclusion is now immediate from the bounded convergence theorem for expectations. $\square$

Laplace functionals have uses similar to those of Laplace transforms:

1.6 PROPOSITION. *Let M and N be random measures on $(E,\mathcal{E})$. They are independent if and only if*

$$\mathbb{E}\ e^{-(Mf+Ng)} = \mathbb{E}\ e^{-Mf}\ \mathbb{E}\ e^{-Ng}, \quad f,g \in \mathcal{E}_+.$$

Proof. To show that M and N are independent, it is sufficient to show that the vectors $(Mf_1,\ldots,Mf_m)$ and $(Ng_1,\ldots,Ng_n)$ are independent for all choices of $m \geq 1$ and $n \geq 1$ and $f_1,\ldots,f_m,g_1,\ldots,g_n$ in $\mathcal{E}_+$. For this, it is enough to show the following equality of joint Laplace transforms:

$$\mathbb{E}\ \exp_-\left(\sum_1^m p_i M f_i + \sum_1^n q_j N g_j\right) = \left(\mathbb{E}\ \exp_- \sum_1^m p_i M f_i\right)\left(\mathbb{E}\ \exp_- \sum_1^n q_j M g_j\right).$$

But this is immediate from the condition of the proposition upon taking $f = \sum p_i f_i$ and $g = \sum q_j g_j$. This proves the sufficiency of the condition. The necessity is obvious. □

The preceding has a useful corollary concerning the Laplace functional of the sum $M+N$; we leave it to Exercise 1.15.

1.7 EXAMPLE. Let λ be a probability measure on $(E,\mathcal{E})$. Let $X = \{X_i : i \in \mathbb{N}^*\}$ be an independency of random variables taking values in $(E,\mathcal{E})$ according to the common distribution λ. Let K be independent of X and have the Poisson distribution with mean c, the last being a constant in $(0,\infty)$. We define M by (recall that $I(x,A) = \delta_x(A) = 1_A(x)$ defines the identity kernel I)

1.8 $$M(\omega,A) = \sum_{i=1}^{K(\omega)} I(X_i(\omega),A), \quad \omega \in \Omega,\ A \in \mathcal{E},$$

where, by the usual conventions, the sum is zero if $K(\omega) = 0$. This defines a measure M_ω for each ω, and re-writing 1.8 as

1.9 $$Mf = \sum_{i=1}^{K} f \circ X_i = \sum_{i=1}^{\infty} f \circ X_i\ 1_{\{K \geq i\}}, \quad f \in \mathcal{E}_+,$$

we see that Mf is a random variable for each f in $\mathcal{E}_+$. Hence M is a random measure. It is integer-valued. Its mean measure is equal to $c\lambda$:

$$\mu f = \mathbb{E}\ Mf = \sum_1^\infty \mathbb{E}\ f \circ X_i\ \mathbb{E}\ 1_{\{K \geq i\}} = (\lambda f)\ \mathbb{E}\ K = c\lambda f.$$

To compute its Laplace functional, we use 1.9 and the assumptions on independence and the distributions of K and the X_i. With the notation e^{-f} for the function $x \mapsto e^{-f(x)}$, we have

$$\mathbb{E}\ e^{-Mf} = \mathbb{E} \prod_{i=1}^{K} e^{-f} \circ X_i = \sum_{k=0}^{\infty} \frac{e^{-c}c^k}{k!}\ (\lambda e^{-f})^k = \exp_-\ c\lambda(1 - e^{-f}).$$

Atoms, point processes

The preceding example is the prototype of the random measures of primary interest, namely, purely atomic random measures.

Let $(\bar{E}, \bar{\mathcal{E}})$ be a measurable space containing $(E, \mathcal{E})$, that is, $\bar{E} \supset E$ and $\bar{\mathcal{E}} \supset \mathcal{E}$. Let $X = (X_i)$ be a countable collection of random variables taking values in $(\bar{E}, \bar{\mathcal{E}})$. Define

1.10 $$Mf(\omega) = \sum_i f \circ X_i(\omega), \quad \omega \in \Omega,\ f \in \mathcal{E}_+,$$

with f extended automatically onto $\bar{E}$ by putting $f(x) = 0$ for all x in $\bar{E} \setminus E$. This defines a random measure M on $(E, \mathcal{E})$. To indicate it, we say that X *forms* M on $(E, \mathcal{E})$ and, conversely, call the X_i *atoms* of M. See Remarks 1.14a and 1.14b below for the need to define X on a larger space $(\bar{E}, \bar{\mathcal{E}})$ in general; often, it is possible to take $E = \bar{E}$.

Assuming that the singleton $\{x\}$ belongs to $\mathcal{E}$ for every x in E, the preceding use of the term "atom" is well-justified: for every ω, the measure M_ω defined by 1.10 is purely atomic, and each $X_i(\omega)$ in E is an atom of M_ω. In particular, if all the $X_i(\omega)$ in E are distinct then M_ω is a counting measure. And if M_ω is such, the simplest way to visualize it is by marking its atoms in E; this yields a set $S_\omega = \{x \in E :\ x = X_i(\omega) \text{ for some } i\}$. The resulting random set S is called the *point process* associated with M.

In practice, often, the random variables X_i are defined first and M is introduced by 1.10. We now treat the converse situation where M is already defined and we need to specify the X_i. The following does this for the case $E = \mathbb{R}_+$ and $\mathcal{E} = \mathcal{B}(\mathbb{R}_+)$; we think of $\mathbb{R}_+$ as time and of atoms as times of occurrence of some physical event. See Exercise 1.19 for the general case.

1.11 PROPOSITION. *Let M be an integer-valued random measure on $\mathbb{R}_+$. Suppose that it is finite over bounded intervals. Then, there exists an increasing sequence (T_n) of random variables taking values in $\bar{\mathbb{R}}_+ = [0, \infty]$ such that, for almost every ω,*

1.12 $$M(\omega, A) = \sum_{n=1}^{\infty} I(T_n(\omega), A), \quad A \in \mathcal{B}(\mathbb{R}_+).$$

Proof. Define $L_t(\omega) = M(\omega, [0, t])$ for t in $\mathbb{R}_+$ and ω in Ω. Since M is a random measure, L_t is a random variable for each t, and by assumption, $L_t < \infty$ almost surely. For each integer $n \geq 1$, let

1.13 $$T_n(\omega) = \inf\{t \in \mathbb{R}_+ :\ L_t(\omega) \geq n\}, \quad \omega \in \Omega,$$

with the usual convention that $\inf \emptyset = +\infty$. Since $\{T_n \leq t\} = \{L_t \geq n\}$ and the latter is an event, T_n is a random variable for each n. Note that T_n takes values in $\bar{\mathbb{R}}_+$ and the sequence (T_n) is increasing. Let Ω_0 be the intersection of $\{L_n < \infty\}$ over all finite integers n; then, Ω_0 is almost sure

since $L_t < \infty$ almost surely for every $t < \infty$. For every ω in Ω_0, the definition 1.13 implies 1.12 for $A = [0, t]$ for every t and, therefore, through a monotone class argument, for every Borel subset A of $\mathbb{R}_+$. □

1.14 REMARKS. a) In the preceding proposition, if one assumes in addition that $M(\mathbb{R}_+) = +\infty$ almost surely, then $T_n < \infty$ almost surely for every n and can be defined to take values in $\mathbb{R}_+$. Otherwise, in general, it is impossible to avoid $+\infty$ as a value for the T_n. For instance, in Example 1.7 with $E = \mathbb{R}_+$, the event $\{K = 3\}$ has a strictly positive probability, and for every ω with $K(\omega) = 3$ we have $0 \le T_1(\omega) \le T_2(\omega) \le T_3(\omega) < +\infty$ and $T_4(\omega) = T_5(\omega) = \cdots = +\infty$. Incidentally, if $K(\omega) = 3$ then $T_1(\omega)$ is the smallest number in $\{X_1(\omega), X_2(\omega), X_3(\omega)\}$, and $T_2(\omega)$ is the second smallest, and $T_3(\omega)$ is the largest.

b) Going back to Example 1.7 on an arbitrary space E, we now re-define the X_n in order that the re-defined sequence $\bar{X}$ form the random measure in the sense of 1.10. Take a point ∂ that is not in E, let $\bar{E} = E \cup \{\partial\}$, and let $\bar{\mathcal{E}}$ be the σ-algebra on $\bar{E}$ generated by $\mathcal{E}$. For each $n \ge 1$, let

$$\bar{X}_n(\omega) = \begin{cases} X_n(\omega) & \text{if } n \le K(\omega), \\ \partial & \text{otherwise.} \end{cases}$$

Then, $\bar{X} = \{\bar{X}_n : n \ge 1\}$ is a sequence of random variables taking values in $(\bar{E}, \bar{\mathcal{E}})$, and $\bar{X}$ forms M by 1.10.

Exercises and complements

1.15 *Sums of independent measures.* Let M and N be independent random measures on $(E, \mathcal{E})$. Show that, then,

$$\mathbb{E}\, e^{-(M+N)f} = \mathbb{E}\, e^{-Mf}\, \mathbb{E}\, e^{-Nf}, \quad f \in \mathcal{E}_+.$$

In words, the Laplace functional of the sum is the product of the Laplace functionals.

1.16 *Mean measure.* Let M be a random measure on $(E, \mathcal{E})$ with mean μ. Suppose that the singleton $\{x\}$ is in $\mathcal{E}$ for every x in E.

a) For f in $\mathcal{E}_+$, if $\mu f < \infty$ then $Mf < \infty$ almost surely.

b) If μ is σ-finite, then M is σ-finite.

c) If x is an atom of μ, then it is an atom of M with a strictly positive probability. If μ is purely atomic, then M is purely atomic.

Show these. The converses of these statements are generally false.

1.17 *Uniform points on $E = [0, 1]$.* Fix an integer $n \ge 1$. Let $X_1, \ldots, X_n$ be independent and uniformly distributed over $E = [0, 1]$. Define M by 1.10 but for Borel functions f on E. Compute the Laplace functional of M.

1.18 *Increasing processes.* Let $L = (L_t)_{t\in\mathbb{R}_+}$ be an increasing right-continuous process with state space $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$. Show that there is a random measure M on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ such that $M(A) = L_t$ for $A = [0, t]$. Hint: Define M_ω to be the measure corresponding to the "distribution function" $t \mapsto L_t(\omega)$.

1.19 *Atoms.* Let $(E, \mathcal{E})$ be a standard measurable space. Let M be a random counting measure with a σ-finite mean measure μ. Let ∂ be a point outside E; put $\bar{E} = E \cup \{\partial\}$ and let $\bar{\mathcal{E}}$ be the σ-algebra on $\bar{E}$ generated by $\mathcal{E}$. Show that there exists a sequence (X_i) of random variables taking values in $(\bar{E}, \bar{\mathcal{E}})$ such that 1.10 holds for M. Hint: Follow the steps of Exercises I.5.15 and I.5.16 to carry M from E to a random measure $\hat{M}$ on $\mathbb{R}_+$, use 1.11 above to pick the atoms $\hat{T}_i$ for $\hat{M}$, and transport the $\hat{T}_i$ back into $\bar{E}$ as the X_i.

1.20 *Lexicographic ordering of atoms.* Let M be a random counting measure on $\mathbb{R}_+ \times F$, where F is a Borel subset of $\mathbb{R}$ (can be replaced with $\mathbb{R}^d$). Suppose that, for almost every ω, the set D_ω of atoms of M_ω has the following pleasant properties:

i) if (s, y) and (t, z) are atoms then $s \neq t$,

ii) the number of atoms in $[0, t] \times F$ is finite for every $t < \infty$, but the total number of atoms in $\mathbb{R}_+ \times F$ is infinite.

For such a good ω, label the atoms $(T_1(\omega), Z_1(\omega)), (T_2(\omega), Z_2(\omega)), \ldots$ going from left to right, so that $0 \le T_1(\omega) < T_2(\omega) < \cdots$. For the remaining negligible event of bad ω, put $T_k(\omega) = 0$ for all $k \ge 1$ and do something similar for the $Z_k(\omega)$. Show that $T_1, T_2, \ldots$ and $Z_1, Z_2, \ldots$ are random variables and that the pairs (T_i, Z_i) form M.

1.21 *Continuation.* Let M be a random counting measure on $\mathbb{R}_+ \times \mathbb{R}$ with mean measure $\mu = \text{Leb} \times \lambda$, where λ is a σ-finite measure on $\mathbb{R}$. Choose a partition (F_n) of $\mathbb{R}$ such that $\lambda(F_n) < \infty$ for every n. Assume that the condition (i) of 1.20 holds, and note that the condition (ii) follows for the restriction of M to $\mathbb{R}_+ \times F_n$ for each n. For n in $\mathbb{N}^*$, let $(T_{n,i}, Z_{n,i})$ with $i = 1, 2, \ldots$ be the lexicographic ordering of the atoms in $\mathbb{R}_+ \times F_n$. Then, the collection $(T, Z) = \{(T_{n,i}, Z_{n,i}) : n \in \mathbb{N}^*, i \in \mathbb{N}^*\}$ forms M.

2 Poisson Random Measures

In this section we introduce Poisson random measures, give some examples, and illustrate their elementary uses.

Recall that a random variable X taking values in $\bar{\mathbb{N}} = \{0, 1, \ldots, \infty\}$ is said to have the Poisson distribution with mean c in $(0, \infty)$ if

2.1
$$\mathbb{P}\{X = k\} = \frac{e^{-c}c^k}{k!}, \quad k \in \mathbb{N},$$

and then $X < \infty$ almost surely and $\mathbb{E}\, X = \text{Var}X = c$. We extend this definition to allow 0 and $+\infty$ as values for c:

2.2 $c = 0 \iff X = 0$ almost surely, $\quad c = +\infty \iff X = +\infty$ almost surely.

Indeed, the case where $c = 0$ is covered by 2.1, and the case where $c = +\infty$ is consistent with 2.1 in a limiting sense.

Recall, also, that if X and Y are independent and have the Poisson distributions with means a and b, then $X + Y$ has the Poisson distribution with mean $c = a + b$. This property extends to countable sums even when the means sum to $c = +\infty$; see Exercise 2.18. These remarks should make it clear that the following definition is without internal contradictions.

2.3 DEFINITION. *Let $(E, \mathcal{E})$ be a measurable space and let ν be a measure on it. A random measure N on $(E, \mathcal{E})$ is said to be* Poisson *with mean ν if*

a) for every A in $\mathcal{E}$, the random variable $N(A)$ has the Poisson distribution with mean $\nu(A)$, and

b) whenever $A_1, \ldots, A_n$ are in $\mathcal{E}$ and disjoint, the random variables $N(A_1), \ldots, N(A_n)$ are independent, this being true for every $n \geq 2$.

2.4 REMARKS. a) In some cases, the condition (a) in the definition implies the condition (b). In fact, the necessary condition that

$$\mathbb{P}\{N(A) = 0\} = e^{-\nu(A)}, \quad A \in \mathcal{E},$$

is also sufficient for N to be Poisson with mean ν at least in the case $E = \mathbb{R}^d$ and $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$ and ν is diffuse, and then it is known that N is also a random counting measure. See Theorem 5.12 for this deep result.

b) *Deterministic transformations.* Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν. Let h be a measurable transformation from $(E, \mathcal{E})$ into another measurable space $(F, \mathcal{F})$. Let M be the image of N under h, that is,

$$M(\omega, B) = N(\omega, h^{-1}B), \quad \omega \in \Omega, \ B \in \mathcal{F};$$

notation: $M = N \circ h^{-1}$. Then, M is a Poisson random measure on $(F, \mathcal{F})$ with mean $\mu = \nu \circ h^{-1}$. This can be shown by checking the conditions in the definition above. In Section 3, this result will be generalized by replacing h with a random transformation.

Examples

2.5 *Particles in boxes.* Let E be countable and $\mathcal{E} = 2^E$, and let ν be a measure on it. For each x in E, let W_x be a Poisson distributed random variable with mean $\nu(\{x\})$. Assume that the countable collection $\{W_x : x \in E\}$ is an independency. Then,

$$N(\omega, A) = \sum_{x \in E} W_x(\omega) I(x, A) \quad \omega \in \Omega, \ A \in \mathcal{E},$$

defines a Poisson random measure N on $(E, \mathcal{E})$ with mean ν. This is the form of the most general Poisson random measure on a space like this. We may think of E as a countable collection of boxes and of W_x as the number of particles in the box x.

2.6 *Stones in a field.* Take a Poisson distributed number of stones. Throw each into a space E, using the same mechanism every time, without regard to the total number of stones or where the previous ones have landed. The final configuration of stones in E describe a Poisson random measure.

The precise version is the construction of Example 1.7. To re-capitulate: let $X = (X_1, X_2, \ldots)$ be an independency of random variables taking values in the measurable space $(E, \mathcal{E})$ and having some probability measure λ as their common distribution. Let K be independent of X and have the Poisson distribution with some number c in $(0, \infty)$ as its mean. For each outcome ω, we think of $K(\omega)$ as the total number of stones and $X_1(\omega), \ldots, X_{K(\omega)}(\omega)$ as the landing points of those stones. The configuration formed by the stones (without regard to their identities) can be described by the measure N_ω where

$$N_\omega(A) = N(\omega, A) = \sum_{n=1}^{K(\omega)} I(X_n(\omega), A), \quad A \in \mathcal{E},$$

that is, the number of stones in the set A. The claim is that the random measure N is Poisson with mean $\nu = c\lambda$. Every Poisson random measure with a finite mean measure can be assumed to have this construction.

To prove the claim, we check the conditions of Definition 2.3. To that end, letting $\{A, \ldots, B\}$ be a finite measurable partition of E, it is enough to show that $N(A), \ldots, N(B)$ are independent and Poisson distributed with respective means $\nu(A), \ldots, \nu(B)$. In view of the assumptions on K and the X_n, we have, for $i, \ldots, j$ in $\mathbb{N}$ with $i + \cdots + j = k$,

$$\begin{aligned} &\mathbb{P}\{N(A) = i, \ldots, N(B) = j\} \\ &\quad = \mathbb{P}\{K = k\}\ \mathbb{P}\{N(A) = i, \ldots, N(B) = j | K = k\} \\ &\quad = \frac{e^{-c}c^k}{k!} \cdot \frac{k!}{i! \cdots j!}\ \lambda(A)^i \cdots \lambda(B)^j = \frac{e^{-\nu(A)}\nu(A)^i}{i!} \cdots \frac{e^{-\nu(B)}\nu(B)^j}{j!}, \end{aligned}$$

which is as needed to be shown.

2.7 *Homogeneous counting measures on the plane.* Let N be a Poisson random measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ with mean $\nu = c\,\text{Leb}$, where c is a constant in $(0, \infty)$ and Leb stands for the Lebesgue measure on $\mathbb{R}^2$. If h is a rigid transformation of the plane, since Lebesgue measure is invariant under such transformations, we have $\nu \circ h^{-1} = \nu$. Thus, in the notation of Remark 2.4b, the image $N \circ h^{-1}$ is again a Poisson random measure with mean ν. In other words, the probability law of N is invariant under rigid transformations; this is the meaning of *homogeneity* for N. Moreover, as will be shown in Theorem 2.17 below, N is a random counting measure, that is, for almost every outcome ω, the measure N_ω is purely atomic with weight one on each atom. In Figure 5 below, the atoms are shown for a typical ω. Poisson random measures are best visualized by means of such pictures. We now give two computations occasioned by such thoughts.

Consider the distance R from the origin to the nearest atom. A glance at the figure shows that $R(\omega) > r$ if and only if $N(\omega, B_r) = 0$, where B_r is

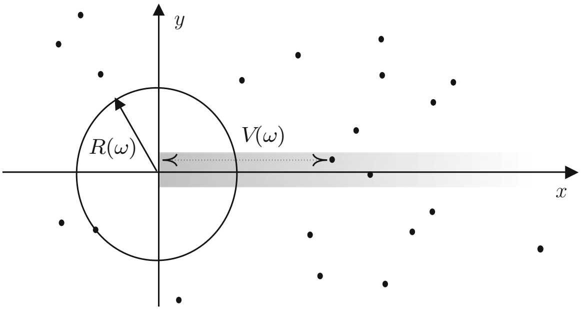


Figure 5: The dots mark the atoms of $N(\omega, \cdot)$, the distance from the origin to the nearest atom is $R(\omega)$, the visibility in the positive x-direction is $V(\omega)$ (the shaded area is $D = [0, \infty) \times [-a, a]$ where a is the "size" of the discs).

the closed disk of radius r centered at the origin. Since $N(B_r)$ is a random variable, $\{R > r\} = \{N(B_r) = 0\}$ is an event; this is true for every r in $\mathbb{R}_+$ and, hence, R is a random variable. Since the area of B_r is πr^2,

$$\mathbb{P}\{R > r\} = \mathbb{P}\{N(B_r) = 0\} = e^{-\nu(B_r)} = e^{-\pi c r^2}, \quad r \in \mathbb{R}_+.$$

For another small computation, we now think of the atoms as centers of small disks of radius a. The disks represent trees in a thin forest with intensity c of trees; assuming that a and c are small, the difficulty of explaining overlapping trees can be ignored. Since N is homogeneous, we may take the positive x-axis as an arbitrary direction. We are interested in the distance V from the origin to the nearest tree intersecting the positive x-axis; this is the visibility in the forest in that direction. Note that a disk intersects the x-axis if and only if its center is at a distance at most a from the x-axis. Thus $V(\omega) > x$ if and only if $N(\omega, D_x) = 0$ where $D_x = [0, x] \times [-a, a]$. It follows that V is a random variable and

$$\mathbb{P}\{V > x\} = \mathbb{P}\{N(D_x) = 0\} = \exp_- \nu(D_x) = e^{-2acx}, \quad x \in \mathbb{R}_+.$$

Mean, variance, Laplace functional

Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν. For A in $\mathcal{E}$, if $\nu(A) < \infty$ then $N(A)$ has a proper Poisson distribution with mean $\nu(A)$ and variance $\nu(A)$. If $\nu(A) = +\infty$ then $N(A) = +\infty$ almost surely, the mean of $N(A)$ is still $\nu(A)$, but the variance is undefined. These carry over to functions as follows: for f in $\mathcal{E}_+$,

2.8 $$\mathbb{E}\, Nf = \nu f, \quad \mathrm{Var} Nf = \nu(f^2) \quad \text{if } \nu f < \infty.$$

The claim for the mean is immediate from 1.3. The one for the variance can be shown by computing $\mathbb{E}\,(Nf)^2$ first for simple f and then for arbitrary f

by approximating f via simple functions. The same steps are used to prove the following important characterization via Laplace functionals.

2.9 THEOREM. *Let N be a random measure on a measurable space $(E, \mathcal{E})$. It is Poisson with mean ν if and only if*

2.10 $$\mathbb{E}\, e^{-Nf} = e^{-\nu(1-e^{-f})}, \quad f \in \mathcal{E}_+.$$

2.11 REMARKS. a) The compact notation on the right side stands for $e^{-\nu g}$ where νg is the integral of the function $g = 1 - e^{-f}$ with respect to the measure ν. In other words, 2.10 can be re-written as

$$\mathbb{E}\, \exp_- \int_E N(dx) f(x) = \exp_- \int_E \nu(dx)(1 - e^{-f(x)}), \quad f \in \mathcal{E}_+.$$

b) Since the Laplace functional of N determines its probability law, we conclude from the theorem above that the probability law of a Poisson random measure is determined by its mean measure.

c) The proof below will show that N is Poisson with mean ν if and only if 2.10 holds for every simple f in $\mathcal{E}_+$.

Proof of Theorem 2.9

Necessity. Suppose that N is Poisson with mean ν. For a in $\mathbb{R}_+$ and A in $\mathcal{E}$ with $\nu(A) < \infty$, since $N(A)$ has the Poisson distribution with mean $\nu(A)$,

$$\mathbb{E}\, \exp_- aN(A) = \sum_0^\infty \frac{e^{-\nu(A)}\nu(A)^k}{k!}\, e^{-ak} = \exp_- \nu(A)(1 - e^{-a});$$

the result remains true even when $\nu(A) = +\infty$. Next, let f in $\mathcal{E}_+$ be simple, say $f = \sum_1^n a_i\, 1_{A_i}$ with the A_i disjoint. Then, $Nf = \sum a_i N(A_i)$ and the variables $N(A_i)$ are independent by the definition of Poisson random measures. So,

$$\mathbb{E}\, e^{-Nf} = \prod_i \mathbb{E}\, \exp_- a_i N(A_i) = \exp_- \sum_i \nu(A_i)(1 - e^{-a_i}),$$

which shows that 2.10 holds when f is simple. Finally, let f in $\mathcal{E}_+$ be arbitrary. Choose $(f_n) \subset \mathcal{E}_+$ increasing to f such that each f_n is simple. By the continuity (Proposition 1.5) of Laplace functionals, using 2.10 for each f_n, we get

$$\mathbb{E}\, e^{-Nf} = \lim_n \mathbb{E}\, \exp_- Nf_n = \lim_n \exp_- \nu(1 - e^{-f_n}).$$

The last limit is equal to the right side of 2.10: as $n \to \infty$, the functions $g_n = 1 - e^{-f_n}$ increase to $g = 1 - e^{-f}$, and the integrals νg_n increase to νg by the monotone convergence theorem.

Sufficiency. This is immediate from the necessity part coupled with the one-to-one relationship between Laplace functionals and probability laws of random measures (Proposition 1.4). □

2.12 Example. *Shot noise.* Arrivals of electrons at an anode form a Poisson random measure N on $\mathbb{R}$ with mean $\nu = c\,\text{Leb}$, where c is a constant in $(0,\infty)$. We view $\mathbb{R}$ as the time axis. Since the Lebesgue measure is diffuse, it follows from Theorem 2.17 below that N is a random counting measure, that is, for almost every ω, no two electrons arrive at the same time. Since the number of electrons arriving during a bounded interval is finite (because the mean is finite), we may assume that the arrival times T_n are ordered so that $\cdots < T_{-1} < T_0 < 0 \le T_1 < T_2 < \cdots$ almost surely.

Each arriving electron produces a current whose intensity is $g(u)$ after a lapse of u time units, and the currents produced by the different electrons are additive. Thus, the resulting current's intensity at time t is

$$X_t = \sum_{n=-\infty}^{\infty} g(t - T_n) 1_{(-\infty,t]} \circ T_n = \int_{(-\infty,t]} N(ds) g(t-s).$$

The function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is generally continuous and decreases rapidly to 0; all we need here is that g be Borel and integrable (over $\mathbb{R}_+$ with respect to the Lebesgue measure). Note that $X_t = Nf$, where for t in $\mathbb{R}$ fixed,

$$f(s) = g(t-s) 1_{(-\infty,t]}(s), \quad s \in \mathbb{R}.$$

Thus, according to 2.8,

$$\begin{aligned} \mathbb{E}\, X_t &= \nu f = c \textstyle\int_{(-\infty,t]} ds\; g(t-s) = c \int_0^\infty du\; g(u), \\ \text{Var} X_t &= \nu(f^2) = c \textstyle\int_{(-\infty,t]} ds\; g(t-s)^2 = c \int_0^\infty du\; g(u)^2, \end{aligned}$$

since the mean is finite by the assumed integrability of g. Finally, we obtain the Laplace transform of X_t by using the preceding theorem on the Laplace functional of N:

$$\begin{aligned} \mathbb{E}\; e^{-rX_t} &= \mathbb{E}\; e^{-N(rf)} \\ &= \exp_- c \int_{(-\infty,t]} ds (1 - e^{-rg(t-s)}) = \exp_- c \int_{\mathbb{R}_+} du (1 - e^{-rg(u)}). \end{aligned}$$

The formulas for the expected value and variance are well-known as *Campbell's theorem.* Variants and generalizations occur frequently. See Exercise 2.25 also.

Finiteness of *Nf*

The following provides a criterion for the almost sure finiteness of the random variable Nf. Recall that $f \wedge g$ is the function whose value at x is the minimum of $f(x)$ and $g(x)$.

2.13 Proposition. *Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν. Let f in $\mathcal{E}_+$ be real-valued.*

a) *If* $\nu(f \wedge 1) < \infty$ *then* $Nf < \infty$ *almost surely.*
b) *If* $\nu(f \wedge 1) = +\infty$ *then* $Nf = +\infty$ *almost surely.*

Proof. We start by recalling that, in view of 2.10 and II.2.31,

2.14 $$\mathbb{P}\{Nf < \infty\} = \lim_{r \to 0} \mathbb{E}\, e^{-rNf} = \lim_{r \to 0} e^{-\nu(1-e^{-rf})}.$$

Moreover, for every $t = f(x)$, the mapping $r \mapsto 1 - e^{-rt}$ is bounded by $t \wedge 1$ on $(0, 1)$ and has the limit 0 at $r = 0$ since $t < \infty$ by the hypothesis that f is real-valued. Thus, as $r \to 0$, the function $1 - e^{-rf}$ is dominated by $f \wedge 1$ and goes to 0. Hence, by the dominated convergence theorem,

$$\nu(f \wedge 1) < \infty \Longrightarrow \lim_{r \to 0} \nu(1 - e^{-rf}) = 0,$$

and this proves the claim (a) via 2.14.

Note that $1 - e^{-t} \geq (1 - e^{-1})(t \wedge 1)$ for $t \geq 0$. This shows, together with the form 2.10 of the Laplace functional, that

$$\nu(f \wedge 1) = +\infty \Longrightarrow \nu(1 - e^{-f}) = +\infty \Longrightarrow \mathbb{E}\, e^{-Nf} = 0,$$

which means that $Nf = \infty$ almost surely, proving the claim (b). □

Existence of Poisson random measures

This is to show that, given a Σ-finite measure ν on a measurable space $(E, \mathcal{E})$, there exists a probability space and a random measure defined over it such that the latter is Poisson with mean ν. The proof is constructive; it is the formal version of the stone throwing we did earlier, repeated a few times.

2.15 Theorem. *Let* ν *be a* Σ*-finite measure on* $(E, \mathcal{E})$. *Then, there exists a probability space* $(W, \mathcal{G}, P)$ *and a measure* $N(w, \cdot)$ *on* $(E, \mathcal{E})$ *for each* w *in* W *such that* N *is Poisson with mean* ν.

Proof. a) First, suppose that ν is finite. Let $c = \nu(E) < \infty$ and define the probability measure μ on $(E, \mathcal{E})$ so that $\nu = c\mu$. Let π be the Poisson distribution on $(\mathbb{N}, 2^{\mathbb{N}})$ with mean c. Define,

$$(W, \mathcal{G}, P) = (\mathbb{N}, 2^{\mathbb{N}}, \pi) \times (E, \mathcal{E}, \mu)^{\mathbb{N}^*};$$

the existence and construction of this follows from the theorem of Ionescu-Tulcea; see Sections 4 and 5 of Chapter IV. Each point w in W is a sequence $w = (x_0, x_1, x_2, \ldots)$; for it, define

$$K(w) = x_0; \quad X_i(w) = x_i, \quad i \in \mathbb{N}^*.$$

Then, $K, X_1, X_2, \ldots$ are totally independent, K has the Poisson distribution π with mean c, and the X_i take values in $(E, \mathcal{E})$ with the distribution μ.

Define N as in Example 2.6 from these variables. As was shown there, then, N is a Poisson random measure on $(E, \mathcal{E})$ with mean $\nu = c\mu$. This completes the proof if ν is finite.

b) Suppose that ν is Σ-finite but not finite. Then, there are finite measures $\nu_1, \nu_2, \ldots$ such that $\nu = \sum \nu_n$. For each n, construct $(W_n, \mathcal{G}_n, P_n)$ and N_n as in the part (a) above, but for the measure ν_n. Now, put

$$(W, \mathcal{G}, P) = \otimes_{n=1}^{\infty} (W_n, \mathcal{G}_n, P_n);$$

see Section 5 of Chapter IV again. For $w = (w_1, w_2, \ldots)$ in W, each w_n is in W_n and we put $\hat{N}_n(w, A) = N_n(w_n, A)$, and finally, define

$$N(w, A) = \sum_{n=1}^{\infty} \hat{N}_n(w, A).$$

Then, $\hat{N}_1, \hat{N}_2, \ldots$ are independent Poisson random measures on $(E, \mathcal{E})$ with mean measures $\nu_1, \nu_2, \ldots$, all defined over the probability space $(W, \mathcal{G}, P)$. Thus, for f in $\mathcal{E}_+$, writing E for the expectation operator corresponding to P,

$$E \exp_- \sum_1^n \hat{N}_i f = \prod_1^n \exp_- \nu_i (1 - e^{-f}) = \exp_- \sum_1^n \nu_i (1 - e^{-f})$$

according to Proposition 1.6 and Theorem 2.9. Letting $n \to \infty$ on both sides we obtain, since $N = \hat{N}_1 + \hat{N}_2 + \cdots$ and $\nu = \nu_1 + \nu_2 + \cdots$,

$$E\, e^{-Nf} = e^{-\nu(1-e^{-f})}, \quad f \in \mathcal{E}_+,$$

which means by Theorem 2.9 that N is Poisson with mean ν as claimed. □

2.16 Remark. *Monte-Carlo.* In Monte-Carlo studies using random measures, the starting point is often the construction of the realization N_ω, for a typical outcome ω, of a Poisson random measure N on a standard measure space. We illustrate the technique for N on $E = \mathbb{R}_+ \times \mathbb{R}_+$ with mean measure $\nu = \text{Leb}$; it is easy to extend the technique to $E = \mathbb{R}^d$. The problem is to get a typical realization of N from a sequence of "random numbers," the latter being the realizations of independent uniform variables over $(0, 1)$. This is a description of the preceding construction in the simulation language, utilizing the σ-finiteness of ν: Pick an appropriately large number a, consider the square $E_0 = (0, a] \times (0, a]$. Generate a Poisson distributed random variable with mean a^2 using the initial random number u_0; if it turns up to be k, use the random numbers $u_1, \ldots, u_{2k}$ to form pairs (au_1, au_2), $(au_3, au_4), \ldots, (au_{2k-1}, au_{2k})$; these k pairs are the atoms, each with unit weight, of a realization of Poisson N_0 on E_0 with $\nu_0 = \text{Leb}$. Repeat this procedure, using fresh random numbers, with the obviously required shifts, to obtain realizations on the squares $E_{ij} = (ia, ja) + E_0$ with i and

j in $\mathbb{N}$; of course, $E_{0,0} = E_0$ and E_{ij} is the translation of E_0 by putting its lower left corner at the point (ia, ja). The resulting collection of atoms in E are the atoms of a typical realization of the Poisson random measure N on $E = \mathbb{R}_+ \times \mathbb{R}_+$ with mean ν equal to the Lebesgue measure.

Poisson counting measures

Let $(E, \mathcal{E})$ be a measurable space and assume that the singleton $\{x\}$ belongs to $\mathcal{E}$ for every x in E. This is the case if $(E, \mathcal{E})$ is a standard measurable space, and, in particular, if E is a Borel subset of some Euclidean space $\mathbb{R}^d$ and $\mathcal{E}$ is the σ-algebra of the Borel subsets of E. The following exploits the preceding construction.

2.17 THEOREM. *Let N be a Poisson random measure on $(E, \mathcal{E})$. Suppose that its mean ν is Σ-finite. Then, N is a random counting measure if and only if ν is diffuse.*

Proof. Necessity. Fix an arbitrary point x in E and let $c = \nu(\{x\})$. Assuming that N is a random counting measure, we need to show that $c = 0$. Indeed, the assumption implies that the event $\{N(\{x\}) \geq 2\}$ is negligible. Whereas, the hypothesis that N is Poisson implies that the same event has probability $1 - e^{-c} - ce^{-c}$. Hence the last probability must vanish, which means that $c = 0$ as needed.

Sufficiency. Assume that ν is diffuse and Σ-finite. Since the probability law of a Poisson random measure is determined by its mean measure, we may assume that N is constructed as in Theorem 2.15; the construction is applicable since ν is Σ-finite. Thus, N has the form

$$N = \sum_1^\infty N_n, \quad N_n = \sum_{i \leq K_n} I(X_{n,i}, \cdot),$$

where the collection $X = \{X_{n,i} : n \geq 1, i \geq 1\}$ is an independency whose every member has a diffuse distribution (since ν is diffuse).

For a pair of distinct indices (n, i) and (m, j), the event $\{X_{n,i} = X_{m,j}\}$ is negligible in view of what was said about X. Thus, the countable union Ω_0 of all such events is negligible. This shows that N is a counting measure, since the form of N_ω is that of a counting measure for every ω outside Ω_0. □

Atomic structure

This is to extend the preceding theorem in a special case of some importance. Consider the case where $E = \mathbb{R}_+ \times \mathbb{R}_+$ and $\mathcal{E} = \mathcal{B}(E)$. For each possible outcome ω, we visualize the atoms of the counting measure N_ω as solid objects, and, if (t, z) in $\mathbb{R}_+ \times \mathbb{R}_+$ is an atom, we create a sense of dynamics by calling t the arrival time of that atom and z its size.

2.18 PROPOSITION. *Let N be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $\nu = \text{Leb} \times \lambda$, where $\lambda\{0\} = 0$ and $\lambda(\varepsilon, \infty) < \infty$ for every $\varepsilon > 0$. Then, for almost every ω, the measure N_ω is a counting measure whose atoms are such that*

a) no atom arrives at time 0, no atom has size 0, no two atoms arrive simultaneously;

b) for every $t < \infty$ and $\varepsilon > 0$, only finitely many atoms arrive before t that have sizes exceeding ε;

c) the preceding statement is true for $\varepsilon = 0$ as well provided that the measure λ be finite; otherwise, during every interval of non-zero length there are infinitely many arrivals of atoms of size at most ε, however small $\varepsilon > 0$ may be.

Proof. a) Almost surely, N puts no mass on $\{0\} \times \mathbb{R}_+$ because $\text{Leb}\{0\} = 0$, and no mass on $\mathbb{R}_+ \times \{0\}$ because $\lambda\{0\} = 0$; let Ω_0 be the almost sure set involved. Fix $\varepsilon > 0$ and consider the random measure M on $\mathbb{R}_+$ defined by $M(A) = N(A \times (\varepsilon, \infty))$, $A \in \mathcal{B}(\mathbb{R}_+)$. Then, M is Poisson with mean $\mu = \lambda(\varepsilon, \infty)\text{Leb}$, and it is a counting random measure by the last theorem. Hence, there is an almost sure set Ω_ε such that, for every ω in it, if (t, z) and (t', z') are atoms of N_ω with $z > \varepsilon$ and $z' > \varepsilon$, then $t \neq t'$. Let Ω_a be the intersection of Ω_0 and all the Ω_ε with $\varepsilon = \frac{1}{2}, \frac{1}{3}, \ldots$; it is almost sure and the statement (a) is true for every ω in it.

b) Since ν puts the mass $t \cdot \lambda(\varepsilon, \infty) < \infty$ on the set $[0, t] \times (\varepsilon, \infty)$, there is an almost sure event $\Omega_{t,\varepsilon}$ on which N has only finitely many atoms in $[0, t] \times (\varepsilon, \infty)$. Let Ω_b be the intersection of $\Omega_{t,\varepsilon}$ over $t = 1, 2, \ldots$ and $\varepsilon = \frac{1}{2}, \frac{1}{3}, \ldots$; it is almost sure, and the statement (b) is true for every ω in it.

c) If λ is finite, let Ω_c be the intersection of $\Omega_{t,0}$ over $t = 1, 2, \ldots$. Otherwise, if λ is an infinite measure, then

$$\nu((t, t + \delta) \times (0, \varepsilon]) = \delta\ \lambda(0, \varepsilon] = +\infty$$

since $\lambda(\varepsilon, \infty) < \infty$ by assumption; this means that there is an almost sure event $\Omega_{t,\delta,\varepsilon}$ on which N has infinitely many atoms in $(t, t + \delta) \times (0, \varepsilon]$; let Ω_c be the intersection of all those almost sure events over $t = 1, 2, \ldots$ and $\varepsilon, \delta = \frac{1}{2}, \frac{1}{3}, \ldots$. The statement (c) is true for every ω in Ω_c.

It is now obvious that the statements (a), (b), (c) hold simultaneously for every ω in the almost sure event $\Omega_a \cap \Omega_b \cap \Omega_c$. □

The preceding proposition will be used in clarifying the jump structure of certain processes constructed from Poisson random measures; see Proposition 4.6 for instance.

Exercises and complements

2.19 *Sums of Poisson variables.* Let $X_1, X_2, \ldots$ be independent random variables having the Poisson distributions with respective means $c_1, c_2, \ldots$. Show

that, then, $X = \sum_1^\infty X_n$ has the Poisson distribution with mean $c = \sum_1^\infty c_n$. Discuss the particular case where all the c_n are finite and $c = +\infty$.

2.20 *Covariances.* Let N be a Poisson random measure with mean ν on some measurable space $(E, \mathcal{E})$. Show that $\mathbb{E}\, N(A)N(B) = \nu(A\cap B)+\nu(A)\nu(B)$ for arbitrary A and B in $\mathcal{E}$; thus, when it exists, covariance of $N(A)$ and $N(B)$ is $\nu(A \cap B)$. Extend these to functions f and g in $\mathcal{E}_+$:

$$\mathbb{E}\, Nf\ Ng = \nu(fg) + \nu f\ \nu g.$$

2.21 *Higher moments.* These can be obtained either directly or via Laplace functionals. For instance, formally, show that

$$\mathbb{E}\,(Nf \cdot Ng)^2 = \lim_{q,r\to 0} \frac{\partial^2}{\partial q^2}\frac{\partial^2}{\partial r^2}\mathbb{E}\, e^{-N(qf+rg)}.$$

2.22 *Product random measure.* Let N be Poisson on $(E, \mathcal{E})$ with mean ν. For each ω, let $M(\omega, \cdot)$ be the product measure $N(\omega, \cdot) \times N(\omega, \cdot)$ on the product space $(E \times E, \mathcal{E} \otimes \mathcal{E})$. Show that M is a random measure whose mean is $\nu \times I + \nu \times \nu$, that is, for every positive h in $\mathcal{E} \otimes \mathcal{E}$,

$$\mathbb{E}\, Mh = \mathbb{E}\int_{E\times E} N(dx)N(dy)h(x,y) = \int_E \nu(dx)h(x,x) + \int_{E\times E} \nu(dx)\nu(dy)h(x,y)$$

Hint: Use 2.20 with $h = 1_{A\times B}$ and a monotone class argument.

2.23 *Arrival processes.* Let N be Poisson on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ with mean $\nu = c$ Leb. Think of N as an arrival process (see Example 2.12) with $\mathbb{R}$ as the time axis, that is, the atoms of N represent the times of arrivals into a store. Let V_t be the length of the interval from t to the first arrival time after t, and let U_t be the amount of time passed since the last arrival before t. Define these carefully. Show that they are random variables; compute

$$\mathbb{P}\{U_t > x, V_t > y\}, \quad x, y \in \mathbb{R}_+,$$

to conclude that U_t and V_t are independent exponential variables.

2.24 *Continuation.* Let N be as in the preceding exercise except that the space now is $\mathbb{R}_+$ instead of $\mathbb{R}$. Define V_t as before, but the definition of U_t needs modification: if $N(\omega, [0,t]) = 0$ let $U_t(\omega) = t$. Re-do the computations of 2.23 for this case. Show that distribution of U_t converges weakly to the exponential distribution as $t \to \infty$.

2.25 *Shot noise.* Let N be as in Example 2.12 but on the space $\mathbb{R}_+$. Let $T_1, T_2, \ldots$ be the successive arrival times after time 0, as before. With $b > 0$ fixed, redefine X_t by

$$X_t = X_0 e^{-bt} + \sum_{n=1}^\infty a e^{-b(t-T_n)} 1_{[0,t]}\circ T_n.$$

a) Assuming that X_0 is independent of (T_n), compute the mean, variance, and Laplace transform of X_t.

b) Show that the distribution of X_t converges weakly, as $t \to \infty$, to the distribution of X_0 in Example 2.12 with $g(t) = a\,e^{-bt}$.

c) Suppose that X_0 here has the limiting distribution found in the preceding part. Show that, then, X_t has the same distribution as X_0 for every t in $\mathbb{R}_+$.

d) Show that (X_t) satisfies the differential equation

$$dX_t = -b\; X_t\; dt + a\; N(dt), \quad t > 0,$$

in other words, for almost every ω,

$$X_t(\omega) = X_0(\omega) - b\int_0^t X_s(\omega)\; ds + aN(\omega, [0,t]).$$

The differential equation shows that this process (X_t) is an *Ornstein-Uhlenbeck* process driven by a Poisson process.

2.26 *Configurations on* $\mathbb{R}^3$. This is similar to Example 2.7. Let N be Poisson on $\mathbb{R}^3$ with mean $\nu = c$ Leb. Again, N is homogeneous and is a random counting measure. Think of its atoms as stars in $\mathbb{R}^3$.

a) Let R be the distance from the origin to the nearest star. Show that it is a random variable. Find its distribution.

b) Let X be the location of the nearest star expressed in spherical coordinates. Find its distribution.

c) Suppose now that each star is a ball of radius a. Let V be the visibility in the x-direction. Find its distribution.

d) Suppose that a camera located at the origin has an angle of vision of 2α radians and is directed in the x-direction (so that the x-axis is the axis of revolution that defines the cone of vision). Let R_α be the distance from the origin to the nearest star within the cone of vision. Find its distribution; note that $R_\pi = R$ in part (a).

2.27 *Conditional structure*. Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν. Let D in $\mathcal{E}$ have $\nu(D) < \infty$. Let $A, \ldots, B$ form a finite measurable partition of D.

a) Show that, for integers $i, \ldots, j$ in $\mathbb{N}$ with $i + \cdots + j = k$,

$$\mathbb{P}\{N(A) = i, \ldots, N(B) = j | N(D) = k\} = \frac{k!}{i! \cdots j!} p^i \cdots q^j,$$

(multinomial distribution) where $p = \nu(A)/\nu(D), \ldots, q = \nu(B)/\nu(D)$. This is another way of saying that, as in Example 2.6, given that $N(D) = k$, the locations of those k stones in D are as if the stones have been thrown into D independently and according to the distribution $\mu(C) = \nu(C)/\nu(D), C \in \mathcal{E} \cap D$.

b) For f in $\mathcal{E}_+$ that vanishes outside D, show that

$$\mathbb{E}(e^{-Nf}|N(D)=k) = [\int_D \mu(dx)e^{-f(x)}]^k, \quad k \in \mathbb{N}.$$

Compute the same conditional expectation for arbitrary f in $\mathcal{E}_+$.

2.28 *Sums.* Let L and M be independent Poisson random measures on $(E, \mathcal{E})$ with means λ and μ. Then, show that $L + M$ is a Poisson random measure with mean $\lambda + \mu$.

2.29 *Continuation.* Let $N_1, N_2, \ldots$ be independent Poisson random measures on $(E, \mathcal{E})$ with means $\nu_1, \nu_2, \ldots$ Show that, then, $N = N_1 + N_2 + \cdots$ is Poisson with mean $\nu = \nu_1 + \nu_2 + \cdots$.

2.30 *Superpositions.* Let F be a countable set and put $\mathcal{F} = 2^F$. Suppose that $\{N_m : m \in F\}$ is an independency of Poisson random measures on $(E, \mathcal{E})$ with mean ν_m for N_m. Define a random measure M on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ by letting

$$M(\omega, A \times \{m\}) = N_m(\omega, A), \quad A \in \mathcal{E},\ m \in F.$$

Show that M is a Poisson random measure and compute its mean μ. We call M the *superposition* of the N_m, $m \in F$, because the atoms of M are pictured by superposing the atoms of the N_m. Make a picture for the case $E = \mathbb{R}_+$, $F = \{1, 2, 3\}$, $\nu_1 = \nu_2 = \nu_3 = \text{Leb}$.

2.31 *Traces.* Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν. Let $D \in \mathcal{E}$. Define

$$\nu_D(B) = \nu(B \cap D), \quad N_D(\omega, B) = N(\omega, B \cap D), \quad B \in \mathcal{E},\ \omega \in \Omega.$$

Then, N_D is called the *trace* of N on D, and ν_D the trace of ν on D. Show that N_D is Poisson on $(E, \mathcal{E})$ with mean ν_D. Show that, if C and D are disjoint sets in $\mathcal{E}$, then N_C and N_D are independent.

2.32 *Singular mean measures.* Let M and N be Poisson random measures on $(E, \mathcal{E})$ with means μ and ν. Suppose that μ and ν are singular with respect to each other (see 5.21 in Chapter I). Show that, then, M and N are independent if $M + N$ is a Poisson random measure.

2.33 *Decomposition into fixed and moving atoms.* Let $(E, \mathcal{E})$ be a standard measurable space; all we need is that $\{x\} \in \mathcal{E}$ for every x in E. Let N be a Poisson random measure on $(E, \mathcal{E})$ with a Σ-finite mean ν. Recall that such ν have at most countably many atoms. Let A be the set of all those atoms and let $D = E \setminus A$. Define ν_A, ν_D, N_A, N_D as the traces as in 2.31. Then,

$$N = N_A + N_D$$

where N_A and N_D are independent Poisson random measures with respective means ν_A and ν_D. Note that N_D is a random counting measure; explain the structure of N_A. Each x in A is an atom of $N(\omega, \cdot)$ for a set of ω with strictly

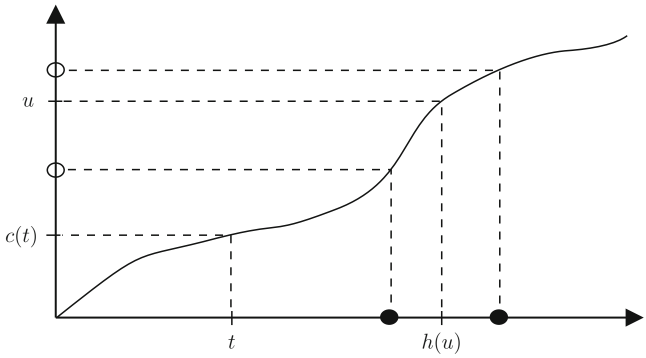


Figure 6: Dots mark the atoms of N, circles mark the atoms of L.

positive probability; such x are called the *fixed* atoms of N. By contrast, the atoms of the counting measure $N_D(\omega, \cdot)$ vary with ω and are called the moving atoms of N.

2.34 *Arrival processes.* Let N be a Poisson random measure on $\mathbb{R}_+$ (with its Borel σ-algebra) with mean ν such that $c(t) = \nu(0, t] < \infty$ for every t in $\mathbb{R}_+$. It is called an arrival process if, in addition, ν is diffuse. Here is a way of constructing such N by means of time changes. Let $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be the functional inverse of the increasing continuous function $t \mapsto c(t)$. Recall that, then, $\nu = \lambda \circ h^{-1}$ where λ is the Lebesgue measure on $\mathbb{R}_+$. Now, let L be a Poisson random measure on $\mathbb{R}_+$ with mean λ, and define $N = L \circ h^{-1}$ in the notation of Remark 2.4b. See Figure 6 above.

2.35 *Intensities.* In the setup of the preceding exercise, suppose that ν is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$, say,

$$\nu(B) = \int_B dt\; r(t), \quad B \text{ Borel},$$

for some positive Borel function $r(t)$, $t \in \mathbb{R}_+$. Then, N is said to have the intensity function r, or $r(t)$ is called the expected arrival rate at time t. Especially when r is bounded, the following technique is effective for constructing N in Monte-Carlo studies.

Let M be a Poisson random measure on $E = \mathbb{R}_+ \times \mathbb{R}_+$ with mean $\mu = \text{Leb}$. Then, for almost every ω, the measure M_ω is a counting measure; let N_ω be the counting measure on $\mathbb{R}_+$ whose atoms are those points t such that (t, z) is an atom of M_ω and $z \le r(t)$. See Figure 7 below.

a) Let M_D be the trace of M on $D = \{(t, z) \in E : z \le r(t)\}$, the last being the region under r. Note that $N = M_D \circ h^{-1}$ where $h : E \mapsto \mathbb{R}_+$ is defined as the projection mapping $(t, z) \mapsto t$. Show that N is Poisson with intensity function r.

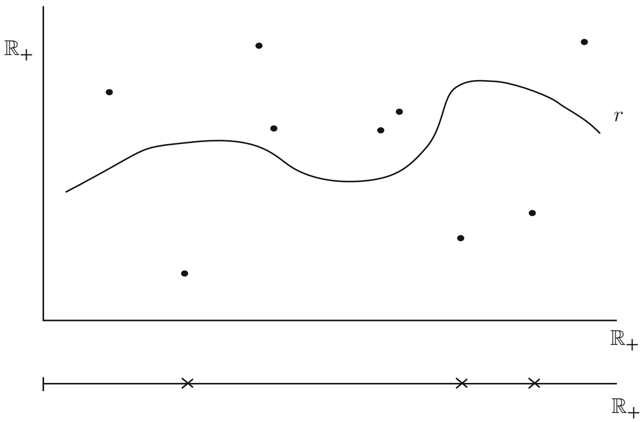


Figure 7: Dots mark the atoms of M on the positive plane. Crosses mark the atoms of N.

b) Show that, for Borel subsets B of $\mathbb{R}_+$, and outcomes ω,

$$N(\omega, B) = \int_{\mathbb{R}_+\times\mathbb{R}_+} M(\omega; dt, dz)\ 1_B(t)\ 1_{[0,r(t)]}(z).$$

Use this to obtain the Laplace functional of N.

2.36 *Random intensities.* Let $R = (R_t)_{t\in\mathbb{R}_+}$ be a bounded positive left-continuous stochastic process. Let M be as in the preceding exercise, a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with unit intensity. Define N from M as before, but with $R_t(\omega)$ replacing $r(t)$ in 2.35b.
Then, N is called an arrival process with the random intensity process R. Generally, N is not Poisson. If R is independent of M, show that

$$\mathbb{E}\ e^{-Nf} = \mathbb{E}\ \exp_- \int_{\mathbb{R}_+} dt\ R_t\ (1 - e^{-f(t)}), \quad f \geq 0 \text{ Borel}.$$

If M and R are independent, then the conditional law of N given R is that of a Poisson random measure on $\mathbb{R}_+$ with intensity function R; such N are said to be conditionally Poisson, or doubly stochastic Poisson, or Cox processes.

2.37 *Conditionally Poisson random measures.* Let L and N be random measures on $(E, \mathcal{E})$. Suppose that the conditional expectation of e^{-Nf} given the σ-algebra $\mathcal{G}$ generated by L has the form

$$\mathbb{E}_{\mathcal{G}} e^{-Nf} = \exp_- \int_E L(dx)(1 - e^{-f(x)}), \quad f \in \mathcal{E}_+.$$

Then, N is said to be conditionally Poisson given L. Preceding exercise is a special case where $E = \mathbb{R}_+$ and $L(dx) = R_x\ dx$.

2.38 *Weak convergence to Poisson laws.* Take n stones. Throw each into the interval $[0, n]$ uniformly at random. For large n, the configuration formed on $\mathbb{R}_+$ is approximately Poisson with unit intensity. Here is the precise version.

Let $U_1, U_2, \ldots$ be independent uniformly distributed random variables on $(0, 1)$. For each integer $n \geq 1$, let M_n be the random measure on $\mathbb{R}_+$ whose atoms are $nU_1, nU_2, \ldots, nU_n$.

a) Show that, for positive Borel f on $\mathbb{R}_+$,

$$\mathbb{E}\, e^{-M_n f} = \left[\frac{1}{n}\int_0^n du\; e^{-f(u)}\right]^n = \left[1 - \frac{1}{n}\int_0^n du\; (1 - e^{-f(u)})\right]^n .$$

b) Assuming that f is continuous, positive, with compact support, note that

$$\lim_{n\to\infty} \mathbb{E}\, e^{-M_n f} = \mathbb{E}\, e^{-Mf},$$

where M is some Poisson random measure on $\mathbb{R}_+$ with unit intensity.

2.39 *Continuation.* In the preceding exercise, the result is that (M_n) "converges in distribution" to a Poisson random measure M with unit intensity. We now explain the meaning of "convergence in distribution" in this context.

Let $\mathcal{M}$ be the collection of all measures on $\mathbb{R}$. Let $C_K = C_K(\mathbb{R} \mapsto \mathbb{R}_+)$ be the collection of all positive continuous functions on $\mathbb{R}$ with compact support. A sequence (μ_n) in $\mathcal{M}$ is said to *converge vaguely* to the measure μ if $\mu_n f \to \mu f$ for every f in C_K. With the topology induced by this mode of convergence, $\mathcal{M}$ becomes a topological space, in fact, a Polish space.

The Borel σ-algebra $\mathcal{B}(\mathcal{M})$ is also the σ-algebra generated by the coordinate functions $\mu \mapsto \mu(A)$ from $\mathcal{M}$ into $\bar{\mathbb{R}}_+$. Thus, we may regard a random measure M as the random variable $\omega \mapsto M_\omega$ taking values in $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$.

Given a sequence (M_n) of random measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, we say that (M_n) *converges in distribution* to the random measure M if

2.40 $$\mathbb{E}\, \varphi \circ M_n \to \mathbb{E}\, \varphi \circ M$$

for every bounded continuous function φ from $\mathcal{M}$ into $\mathbb{R}$. This is the natural generalization of the concept discussed in Chapter III to the space $\mathcal{M}$. In fact, it is sufficient to check 2.40 for functions φ of the form $\varphi(\mu) = e^{-\mu f}$ with f in C_K. We state this here without proof. The following statements are equivalent:

a) (M_n) converges in distribution to M.
b) $(M_n f)$ converges in distribution to Mf for every f in C_K.
c) $(\mathbb{E}\; \exp_\; M_n f)$ converges to $\mathbb{E}\; \exp_\; Mf$ for every f in C_K.

3 Transformations

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Let $X = \{X_i : i \in I\}$ and $Y = \{Y_i : i \in I\}$ be collections, indexed by the same countable set I, of random variables taking values in $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively.

Suppose that X forms a Poisson random measure on $(E, \mathcal{E})$, that is, the random measure N on $(E, \mathcal{E})$ defined by

3.1 $$Nf = \sum_{i \in I} f \circ X_i, \quad f \in \mathcal{E}_+,$$

is Poisson with some mean measure ν. Suppose also that, for some measurable transformation $h : E \mapsto F$, we have $Y_i = h \circ X_i$ for each i. Then, the random measure formed by Y on $(F, \mathcal{F})$ is the image $N \circ h^{-1}$ of N under h and is again Poisson; see Remark 2.4b. In this section we consider two generalizations of this: first, we let each Y_i be a random transform of the corresponding X_i according to the heuristic that Y_i falls in B with probability $Q(x, B)$ if $X_i = x$. Second, we regard each Y_i as an indicator of some property associated with the atom X_i, which leads us to the random measure M formed by $(X, Y) = \{(X_i, Y_i) : i \in I\}$ as a marked version of N.

We present the main ideas in terms of the setup above, which is convenient in most applications. Afterward, in Theorem 3.19, we handle the more general case where the atoms X_i and Y_i take values in spaces larger than E and F. Finally, in Theorem 3.26, we give a modern re-formulation of the main results directly in terms of counting measures.

Main theorem

This is important; used with some art, it simplifies many a complex problem to mere computations.

3.2 THEOREM. *Let ν be a measure on $(E, \mathcal{E})$, and Q a transition probability kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Assume that (i) the collection X forms a Poisson random measure with mean ν, and (ii), given X, the variables Y_i are conditionally independent and have the respective distributions $Q(X_i, \cdot)$. Then,*

a) Y forms a Poisson random measure on $(F, \mathcal{F})$ with mean νQ, and

b) (X, Y) forms a Poisson random measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\nu \times Q$.

3.3 REMARK. Recall from I.6.23 that $\mu = \nu \times Q$ means that

$$\mu(dx, dy) = \nu(dx) Q(x, dy),$$

and νQ is the marginal of μ on F.

Proof. Let N be the random measure formed by X on $(E, \mathcal{E})$, and M the one formed by (X, Y) on $(E \times F, \mathcal{E} \otimes \mathcal{F})$. Note that the random measure formed by Y is the image of M under the projection mapping $h(x, y) = y$. Thus, (a) is immediate from (b), and we shall prove (b) by showing that the Laplace functional of M has the form required by Theorem 2.9. Note that, for positive real-valued f in $\mathcal{E} \otimes \mathcal{F}$,

$$e^{-Mf} = \prod_{i \in I} e^{-f} \circ (X_i, Y_i).$$

In view of the assumption (ii), the conditional expectation of e^{-Mf} given X is equal to

$$\prod_i \int_F Q(X_i, dy) e^{-f} \circ (X_i, y) = \prod_i e^{-g} \circ X_i = e^{-Ng},$$

where g is defined by

$$e^{-g(x)} = \int_F Q(x, dy) e^{-f(x,y)}.$$

It follows that

3.4 $$\mathbb{E}\, e^{-Mf} = \mathbb{E}\, e^{-Ng} = \exp_\ \nu(1 - e^{-g})$$

where we used Theorem 2.9 on the Laplace functional of N after noting that, by the assumption (i), N is Poisson with mean ν. Since $Q(x, F) = 1$,

$$\nu(1 - e^{-g}) = \int_E \nu(dx) \int_F Q(x, dy)(1 - e^{-f(x,y)}) = (\nu \times Q)(1 - e^{-f}),$$

and, putting this into 3.4, we conclude from Theorem 2.9 that M is Poisson with mean $\nu \times Q$ as claimed. □

The following is an immediate corollary where $Q(x, B)$ is specialized to become $\pi(B)$ for some probability measure π. No proof is needed.

3.5 Corollary. *Suppose that X forms a Poisson random measure on $(E, \mathcal{E})$ with mean ν, and that Y is independent of X and is an independency of variables with distribution π on $(F, \mathcal{F})$. Then, (X, Y) forms a Poisson random measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\nu \times \pi$.*

In certain applications, it is convenient to think of Y_i as some mark (weight, velocity, etc.) associated with the atom X_i. Then, (X, Y) is sometimes called a *marked point process* on E with mark space F, and we may think of the Poisson random measure on $E \times F$ as a magnification of that on E. The next four sub-sections provide examples on the uses of the theorem and corollary above.

Compound Poisson processes

The arrival times T_i of customers at a store form a Poisson random measure N on $\mathbb{R}_+$ with intensity c, that is, the mean measure is $\nu = c\text{Leb}$. The customers spend, independently of each other, random amounts of money at the store, the mean being a, variance b^2, and distribution π. We are interested in Z_t, the cumulative amount spent by all who arrived at or before t.

More precisely, we are assuming that the customer who arrives at T_i spends an amount Y_i, where $Y = (Y_i)$ is independent of $T = (T_i)$ and the

variables Y_i are independent and have the distribution π on $\mathbb{R}_+$ in common (with mean a and variance b^2). It follows from the preceding corollary that (T, Y) forms a Poisson random measure M on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $\mu = \nu \times \pi$. For fixed t, the variable Z_t is defined by M through

3.6 $$Z_t = \sum_{i=1}^{\infty} Y_i \; 1_{(0,t]} \circ T_i = \int_{[0,t] \times \mathbb{R}_+} M(ds, dy) y = Mf,$$

where $f(s, y) = y \; 1_{[0,t]}(s)$. It follows from 2.8 and Theorem 2.9 applied to M with this f that

3.7 $$\mathbb{E} \; Z_t = \mu f = act, \quad \text{Var} \; Z_t = \mu(f^2) = (a^2 + b^2)ct$$
$$\mathbb{E} \; \exp_ \; rZ_t = \mathbb{E} \; e^{-M(rf)} = \exp_ \; ct \int_{\mathbb{R}_+} \pi(dy)(1 - e^{-ry}).$$

The process $Z = (Z_t)_{t \in \mathbb{R}_+}$ is an example of compound Poisson processes. The most general cases are obtained by allowing the Y_i to be $\mathbb{R}^d$-valued, without restrictions on the distribution π on $\mathbb{R}^d$, with the same assumptions of independence.

Money in the bank

This is the same as the shot noise process of Example 2.12, but the deterministic function g is replaced by a randomized one. Let $T_1, T_2, \ldots$ form a Poisson random measure N on $\mathbb{R}_+$ with mean $\nu = a$ Leb. We think of T_i as the arrival time of the ith person to a bank in order to open an account; let $Y_i(u)$ be the balance at time $T_i + u$ for that account. Then, the sum of all balances at time t is (assuming $X_0 = 0$)

3.8 $$X_t = \sum_{i=1}^{\infty} Y_i(t - T_i) \; 1_{[0,t]} \circ T_i.$$

We suppose that the processes $Y_1, Y_2, \ldots$ are independent of each other and of the collection (T_i), and let

$$P_u(B) = \mathbb{P}\{Y_i(u) \in B\}, \quad u \in \mathbb{R}_+, B \in \mathcal{B}(\mathbb{R}_+).$$

We are interested in the mean, variance, and Laplace transform of X_t for fixed t. Exercise 2.25 is the special case where $Y_i(u) = g(u)$ for all i and u.

We assume that the Y_i are right-continuous and bounded. Thus, each Y_i takes values in the space F of all bounded right-continuous functions $y : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $\mathcal{F}$ the Borel σ-algebra corresponding to the supremum norm. Since (T_i) forms the Poisson random measure N on $\mathbb{R}_+$, it follows from Corollary 3.5 that the pairs (T_i, Y_i) form a Poisson random measure M on $(\mathbb{R}_+ \times F, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{F})$ with mean $\mu = \nu \times \pi$, where $\nu = a$ Leb and π is the probability law of Y_i. The law π is not specified, but we are given

$$\pi\{y \in F : y(u) \in B\} = P_u(B).$$

Finally, we note that 3.8 is the same as

3.9 $$X_t = Mf, \quad \text{where } f(s,y) = y(t-s)\ 1_{[0,t]}(s), \quad s \in \mathbb{R}_+,\ y \in F.$$

We leave the mean and variance as an exercise to compute and do the Laplace transform of X_t for fixed t by using Theorem 2.9 on M and f here: for r in $\mathbb{R}_+$,

3.10 $$\begin{aligned}\mathbb{E}\ e^{-rX_t} &= \mathbb{E}\ e^{-M(rf)} = e^{-\mu(1-e^{-rf})} \\ &= \exp_\ a\int_{[0,t]} ds \int_F \pi(dy)\ (1-e^{-ry(t-s)}) \\ &= \exp_\ a\int_{[0,t]} ds\ \mathbb{E}(1-e^{-rY_i(t-s)}) \\ &= \exp_\ a\int_{[0,t]} du \int_{\mathbb{R}_+} P_u(dx)(1-e^{-rx}).\end{aligned}$$

Closed particle systems

Imagine some particles moving about in space E. At time 0, the configuration of particles form a Poisson random measure N_0 on $(E, \mathcal{E})$ with some mean ν_0. Each particle moves in E according to a probability law P^x if its initial position is x. Other than this dependence on the initial positions, the particle motions are independent. We are interested, for fixed time t, in the random measure N_t formed by the positions of the particles at time t.

To make the picture more precise, we label the particles with the integers $i \geq 1$, let X_i be the initial position of the particle i, and let $Y_i = (Y_i(t))_{t\in\mathbb{R}_+}$ be the corresponding motion with $Y_i(0) = X_i$. Each Y_i is a stochastic process with state space $(E, \mathcal{E})$; we view it as a random variable taking values in the function space $(F, \mathcal{F}) = (E, \mathcal{E})^{\mathbb{R}_+}$. We are given P^x as the conditional distribution of Y_i given that $X_i = x$, and the X_i form a Poisson random measure N_0 on $(E, \mathcal{E})$ with mean ν_0, and we may assume that $Q(x, B) = P^x(B)$ defines a transition kernel. Then, Theorem 3.2 applies, and the Y_i form a Poisson random measure M on $(F, \mathcal{F})$ with mean $\mu = \nu_0 Q$, that is, with

$$\mu(B) = \int_E \nu_0(dx) P^x(B), \quad B \in \mathcal{F}.$$

Most everything about this particle system can be posed in terms of M. In particular, $N_t = M \circ h^{-1}$ where $h(w) = w(t)$ for w in F and t fixed. Thus, N_t is a Poisson random measure on $(E, \mathcal{E})$ with mean $\nu_t = \mu \circ h^{-1}$, that is,

3.11 $$\nu_t(A) = \int_E \nu_0(dx) \int_F P^x(dw) 1_A(w(t)) = \int_E \nu_0(dx) P_t(x, A), \quad A \in \mathcal{E},$$

where

3.12 $$P_t(x, A) = P^x\{w \in F : w(t) \in A\}$$

is the probability that a particle started at x is in A at time t. Indeed, this could have been obtained more directly by noting that N_t is formed by the atoms $Y_i(t)$, each of which has the conditional distribution $P_t(x, \cdot)$ given that $Y_i(0) = X_i = x$.

Here is a particular case of interest. Take $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$, $\nu_0 = \text{Leb}$, and assume that every particle motion is Brownian (independent of all others). Then,

3.13 $$P_t(x, A) = \int_A dz \, \frac{1}{\sqrt{2\pi t}} e^{-(z-x)^2/2t}, \quad A \in \mathcal{E},$$

and we observe that $\nu_t = \nu_0 P_t = \nu_0 = \text{Leb}$. Thus, in this case, the particle configuration at time t is Poisson with mean $\nu_t = \text{Leb}$ for all times t. In this sense, the particle system is in equilibrium, even though individual particles never experience equilibrium.

In the theory of Markov processes on general state spaces $(E, \mathcal{E})$, the family $(P_t)_{t \in \mathbb{R}_+}$ of transition kernels is called the *transition semigroup*, and measures ν satisfying $\nu = \nu P_t$ are said to be *invariant*. When ν is an infinite invariant measure, the particle system above with $\nu_0 = \nu$ provides a dynamic meaning for ν.

Particle systems with birth and death

Particles arrive over time according to a Poisson process on $\mathbb{R}$ with intensity a. Each arriving particle lands somewhere in E and starts moving in E in some random fashion until it dies. Thus, at each time, a snapshot of E will show the locations of the particles that are alive (born but not dead) at that time. We are interested in the evolution of this picture over time.

We label the particles with integers i in $\mathbb{Z}$, let T_i be the arrival time for i, and Y_i its "motion". Each Y_i is a stochastic process $(Y_i(t))_{t \in \mathbb{R}_+}$ with state space $(\bar{E}, \bar{\mathcal{E}})$, where $\bar{E} = E \cup \{\partial\}$ and $\bar{\mathcal{E}}$ is the σ-algebra on $\bar{E}$ generated by $\mathcal{E}$. We regard ∂ as the cemetery attached to E; it is a trap; once a particle enters ∂ it must stay there forever. We interpret $Y_i(t)$ as the location of i at time $T_i + t$, being in state ∂ means being dead. We regard Y_i, $i \in \mathbb{Z}$, as random variables taking values in the function space $(F, \mathcal{F}) = (\bar{E}, \bar{\mathcal{E}})^{\mathbb{R}_+}$; they are assumed to be independent of each other and of the T_i, and we let P their common probability law. It follows from Corollary 3.5 that the pairs (T_i, Y_i) form a Poisson random measure M on $(\mathbb{R} \times F, \mathcal{B}_{\mathbb{R}} \otimes \mathcal{F})$ with mean $\mu = a \, \text{Leb} \times P$.

Consider the snapshot at time t; it shows the locations $Y_i(t - T_i)$ of those particles i that are born but not dead, that is, $T_i \le t$ and the location

not ∂. The snapshot can be represented by the random measure M_t on $(E, \mathcal{E})$ defined by

3.14 $$M_t(A) = \sum_{i=-\infty}^{\infty} 1_A \circ Y_i(t - T_i)\ 1_{(-\infty,t]} \circ T_i, \quad A \in \mathcal{E}.$$

Essentially, M_t is the trace on E of $\bar{M}_t = M \circ h^{-1}$, where $h : \mathbb{R} \times F \mapsto E$ is given by

$$h(s, w) = \begin{cases} w(t-s) & \text{if } s \le t, \\ \partial & \text{if } s > t. \end{cases}$$

Since M is Poisson, $\bar{M}_t$ is Poisson on $\bar{E}$ by Remark 2.4b, and M_t is Poisson on E by Exercise 2.31 on traces. As to the mean measure μ_t of M_t, we have, for $A \in \mathcal{E}$,

3.15 $$\begin{aligned} \mu_t(A) &= a \int_{(-\infty,t]} ds \int_F P(dw)\ 1_A \circ w(t-s) \\ &= a \int_{\mathbb{R}_+} du\ \mathbb{E}\ 1_A \circ Y_0(u) = a\ \mathbb{E} \int_{\mathbb{R}_+} du\ 1_A \circ Y_0(u) \end{aligned}$$

In summary, M_t is a Poisson random measure on $(E, \mathcal{E})$, and the mean number of particles in A at anytime t is equal to the arrival rate a times the expected amount of time spent in A by one particle during its lifetime.

The computations can be made more specific by making the law P of the Y_i more precise. We illustrate this by assuming that the Y_i are independent replicas of the process $X = (X_t)_{t \in \mathbb{R}_+}$ with state space $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$, defined by

3.16 $$X_t(\omega) = \begin{cases} X_0(\omega) + W_t(\omega) & \text{if } t < \zeta(\omega), \\ +\infty & \text{if } t \ge \zeta(\omega), \end{cases}$$

where X_0, W, ζ are independent, X_0 has the Gaussian distribution with mean 0 and variance b, and W is a Wiener process, and ζ has the exponential distribution with parameter c. Note that X describes the motion of a particle that starts at X_0, moves as a standard Brownian motion, and dies at age ζ and is carried to the point $\partial = +\infty$. In this case, 3.15 becomes, for $A \in \mathcal{B}_{\mathbb{R}}$,

3.17 $$\begin{aligned} \mu_t(A) &= a \int_{\mathbb{R}_+} du\ \mathbb{P}\{X_u \in A\} \\ &= \frac{a}{c} \int_{\mathbb{R}_+} du\ ce^{-cu} \int_{\mathbb{R}} \pi(dx) P_u(x, A) \end{aligned}$$

where π is the distribution of X_0, and $P_u(x, A)$ is as in 3.13. The integral over $\mathbb{R}$ is the probability that $X_0 + W_u$ belongs to A. Thus, the double integral yields the probability that $X_0 + W_\zeta$ is in A. And, we know by calculating the characteristic function of W_ζ that W_ζ has the same distribution as $Z_1 - Z_2$, where Z_1 and Z_2 are independent and exponentially distributed random

variables with parameter $\sqrt{2c}$. Letting ν be the distribution of $X_0 + Z_1 - Z_2$, we see that 3.17 is the same as

3.18 $$\mu_t(A) = \frac{a}{c}\,\nu(A).$$

In particular, we note that the total number of particles in E has the Poisson distribution with mean $\mu_t(E) = \frac{a}{c}$, the expected number of new arrivals during one lifetime.

Generalization of the main theorem

The setup and assumptions of the main theorem, 3.2, include two points: the index set I is countable, and X forms a Poisson random measure with mean ν. Note that, being deterministic, I has to be infinite, and thus, ν must be an infinite measure. These conditions are caused by letting the X_i take values in E. To remedy the situation, we let them take values in a larger space $\bar{E}$ as in the discussion on atoms; see 1.10 *et seq.* and Remark 1.14b.

Let $(\bar{E}, \bar{\mathcal{E}})$ be a measurable space that contains $(E, \mathcal{E})$, let $(\bar{F}, \bar{\mathcal{F}})$ similarly contain $(F, \mathcal{F})$ by setting $\bar{F} = F \cup \{\Delta\}$ with an extra point Δ outside F. Let I be a countably infinite index set as before, $X = \{X_i : i \in I\}$ a collection of random variables taking values in $(\bar{E}, \bar{\mathcal{E}})$, and $Y = \{Y_i : i \in I\}$ in $(\bar{F}, \bar{\mathcal{F}})$. Every function on E is extended onto $\bar{E}$ by letting it vanish on $\bar{E} \setminus E$, and similarly for extension from F onto $\bar{F}$ and from $E \times F$ onto $\bar{E} \times \bar{F}$. Given a transition probability kernel Q from $(E, \mathcal{E})$ into $(F, \mathcal{F})$, we extend it to a kernel $\bar{Q}$ from $(\bar{E}, \bar{\mathcal{E}})$ into $(\bar{F}, \bar{\mathcal{F}})$ by the requirement that $\bar{Q}(x, F) = 0$ and $\bar{Q}(x, \bar{F}) = 1$ for x in $\bar{E} \setminus E$. Finally, recall the meaning of "X forms a random measure N on $(E, \mathcal{E})$", namely, that $Nf = \sum_i f \circ X_i$ for f in $\mathcal{E}_+$ extended onto a function on $\bar{\mathcal{E}}_+$ as prescribed. With this setup, the following is the generalization of Theorem 3.2.

3.19 THEOREM. *Suppose that X forms a Poisson random measure on $(E, \mathcal{E})$ with mean ν. Assume that, given X, the Y_i are conditionally independent with corresponding distributions $\bar{Q}(X_i, \cdot)$. Then, Y forms a Poisson random measure on $(F, \mathcal{F})$ with mean νQ, and (X, Y) forms a Poisson random measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\nu \times Q$.*

Proof. This follows the proof of 3.2 word for word except for the substitution of $\bar{Q}$ for Q in some places. □

The preceding is the most general result on transformations of Poisson random measures. Unfortunately, its formulation is in terms of the atoms rather than being directly in terms of the random measures. The following is aimed at this direct formulation.

Random transformations of Poisson

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. By a *random transformation* from E into F we mean a mapping

3.20 $$\varphi : \quad (\omega, x) \mapsto \varphi(\omega, x)$$

that is measurable relative to $\mathcal{H} \otimes \mathcal{E}$ and $\mathcal{F}$. We write φx for the random variable $\omega \mapsto \varphi(\omega, x)$ and φ_ω for the transformation $x \mapsto \varphi_\omega x = \varphi(\omega, x)$. Of course, φ can be regarded as a collection $\varphi = \{\varphi x : x \in E\}$, and it is said to be an independency if the collection is such in the usual sense. In that case, the probability law of φ is specified by the marginal distributions

3.21 $$Q(x, B) = \mathbb{P}\{\varphi x \in B\}, \quad x \in E, \quad B \in \mathcal{F}.$$

The joint measurability assumed for the mapping 3.20 assures that the preceding defines a transition probability kernel Q from $(E, \mathcal{E})$ into $(F, \mathcal{F})$.

Given a random measure N on $(E, \mathcal{E})$ and a random transformation φ from $(E, \mathcal{E})$ into $(F, \mathcal{F})$, we define the image of N under φ as the random measure $\hat{N}$ on $(F, \mathcal{F})$, and write $N \circ \varphi^{-1}$ for $\hat{N}$, defined by

3.22 $$\hat{N}_\omega f = (N_\omega \circ \varphi_\omega^{-1}) f = \int_E N(\omega, dx)\ f(\varphi_\omega x), \quad \omega \in \Omega,\ f \in \mathcal{F}_+.$$

Magnification M of N is defined similarly

3.23 $$M_\omega f = \int_E N(\omega, dx)\ f(x, \varphi_\omega x), \quad \omega \in \Omega,\ f \in (\mathcal{E} \otimes \mathcal{F})_+.$$

Assuming that N and φ are independent, with ν as the mean of N and Q as in 3.21, we observe, by conditioning on φ first and using Fubini's theorem repeatedly, that

3.24 $$\mathbb{E}\ \hat{N} f = \mathbb{E} \int_E \nu(dx) f(\varphi x) = \int_E \nu(dx) \int_F Q(x, dy) f(y) = \nu Q f,\ f \in \hat{\mathcal{F}},$$

3.25 $$\mathbb{E}\ M f = \mathbb{E} \int_E \nu(dx) f(x, \varphi x)$$
$$= \int_E \nu(dx) \int_F Q(x, dy) f(x, y) = (\nu \times Q) f, \quad f \in \mathcal{E} \otimes \mathcal{F},\ f \geq 0.$$

The following is the promised direct formulation.

3.26 THEOREM. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be standard measurable spaces. Let ν be a σ-finite diffuse measure on $(E, \mathcal{E})$, and Q a transition probability kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Suppose that N is Poisson on $(E, \mathcal{E})$ with mean ν, and that φ is independent of N and is an independency of variables φx with distributions 3.21. Then, $\hat{N}$ is Poisson on $(F, \mathcal{F})$ with mean νQ, and M is Poisson on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\nu \times Q$.*

3.27 REMARKS. a) The conditions of the theorem imply some side conclusions: N exists and is a random counting measure by Theorems 2.15 and 2.16; and there exist random variables $X_1, X_2, \ldots$ taking values in $(\bar{E}, \bar{\mathcal{E}})$ with $\bar{E} = E \cup \{\partial\}$ such that N is formed by them, this is by Exercise 1.18. Moreover, since $(F, \mathcal{F})$ is standard, Kolmogorov's extension theorem IV.4.18 applies to show that φ exists as an independency indexed by E.

b) The condition that ν be diffuse is to ensure that N be a random counting measure. Otherwise, the claims of the theorem are false. The reason is that, if ν has an atom x, then N will have a (Poisson distributed) number of stones at x and all those stones will be transferred to the same random point φx in F . This makes $\hat{N}$ not Poisson.

Proof. Let $\bar{E} = E \cup \{\partial\}$ and $\bar{F} = F \cup \{\Delta\}$, and let $\bar{\mathcal{E}}$ and $\bar{\mathcal{F}}$ be the σ-algebras on $\bar{E}$ and $\bar{F}$ respectively, generated by $\mathcal{E}$ and $\mathcal{F}$ respectively. As remarked in 3.27a, there exist random variables $X_1, X_2, \ldots$ taking values in $(\bar{E}, \bar{\mathcal{E}})$ that form N on $(E, \mathcal{E})$. Define Y_n by setting $Y_n(\omega) = \varphi(\omega, X_n(\omega))$ after extending φ onto $\bar{E}$ by letting $\varphi(\omega, \partial) = \Delta$ for all ω. The joint measurability of $(\omega, x) \mapsto \varphi(\omega, x)$ ensures that $Y_1, Y_2, \ldots$ are random variables taking values in $(\bar{F}, \bar{\mathcal{F}})$. Now, (X, Y) satisfy all the conditions of Theorem 3.19, and the proof is immediate. □

Exercises and complements

3.28 *Heuristics.* This is to give an informal "proof" of Theorem 3.19 at least for the case where ν is finite. Take a Poisson distributed number of stones with mean c. Throw each into E as in Example 2.6; the resulting configuration is Poisson with mean measure $\nu = c\lambda$. Next, take each stone in E and throw into F, independently of all others, so that the stone at the point x of E lands in the subset B of F with probability $Q(x, B)$. The resulting configuration of stones in F must form a Poisson random measure with mean νQ, because the net effect of all the stone throwing is that a Poisson distributed number of stones with mean c got thrown into F according to the distribution λQ. Replacing F by $E \times F$ we also get the marking result.

3.29 *Marked point processes.* Let X and Y be as in the setup leading to Theorem 3.2. Consider the atoms X_i as the points of a point process on E, and regard each Y_i as a *mark* associated with the corresponding atom X_i. Then, some authors refer to (X, Y) as a *marked point* process on E with mark space F.

3.30 *Random decompositions.* This is the converse to the superposition described in Exercise 2.29. Let $X = \{X_i : i \in I\}$ form a Poisson random measure N on $(E, \mathcal{E})$ with mean ν. Suppose that each atom X_i has a mark Y_i, the latter being m with probability $p_m(x)$ if the atom is at x, where $x \in E$ and $m \in F = \{1, 2, \ldots\}$. Let N_m be the random measure on $(E, \mathcal{E})$ formed by the atoms of N marked m. Formulate this story in precise terms

and supply the missing assumptions. Show that $N_1, N_2, \ldots$ are independent Poisson random measures with means $\nu_1, \nu_2, \ldots$ respectively, where

$$\nu_m(dx) = \nu(dx)p_m(x), \quad x \in E,\ m \in F.$$

3.31 *Continuation.* Arrivals of male and female customers at a store form independent Poisson random measures N_1 and N_2 on $\mathbb{R}_+$ with intensities a and b respectively, that is, the mean of N_1 is $\nu_1 = a$ Leb, and of N_2 is $\nu_2 = b$ Leb. What is the probability that exactly 5 males arrive during the interval from 0 to the time of first female arrival. Answer: $p^5(1-p)$, where $p = a/(a+b)$, obviously!

3.32 *Translations.* Let X and Y be as in the setup preceding 3.2, but with $E = F = \mathbb{R}^d$. Suppose that the conditions of Corollary 3.5 are satisfied. Show that, then, $\{X_i + Y_i : i \in I\}$ forms a Poisson random measure $\hat{N}$ on $\mathbb{R}^d$ with mean $\hat{\nu} = \nu * \pi$, the last being the convolution of ν and π, that is,

$$\hat{\nu} f = \int_E \nu(dx) \int_F \pi(dy) f(x+y).$$

Show that, if $\nu = c$ Leb for some constant c, then $\hat{\nu} = \nu$. Much of the sub-section on closed particle systems can be reduced to this case.

3.33 *Particle systems with birth and death.* In the setup of the corresponding subsection, suppose that the law P of the Y_i gives

$$\pi_t(A) = P\{w : w(t) \in A\} = \mathbb{P}\{Y_0(t) \in A\}, \quad A \in \mathcal{E},$$

for $t \in \mathbb{R}_+$. Here π_t is a defective probability measure, the defect $1 - \pi_t(E)$ being the probability that the particle died before t. Let $f \in \mathcal{E}_+$, and interpret $f(x)$ as the rate, per unit time, at which a particle pays "rent" when its position is x; of course, $f(\partial) = 0$ extends f onto $\bar{E}$. Then,

$$W_t = \sum_i \int_0^t ds\ e^{-rs}\ f(Y_i(s - T_i))\ 1_{(-\infty,s]} \circ T_i$$

is the total rent paid by all the particles during $[0,t]$, discounted at rate r. Compute the expected value and variance of W_t and of W_∞.

3.34 *Compound Poisson random measures.* Let X and Y satisfy the setup and conditions of Corollary 3.5. Then, (X, Y) forms a Poisson random measure M on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\mu = \nu \times \pi$. Now, take $F = \mathbb{R}_+$ and define

$$L(A) = \int_{A \times F} M(dx, dy)\ y, \quad A \in \mathcal{E}.$$

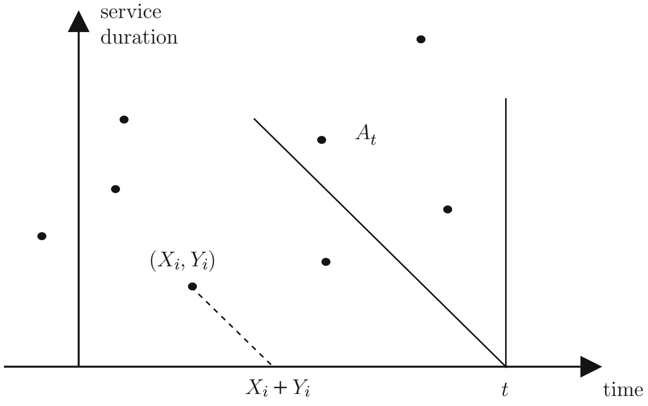


Figure 8: The pairs (X_i, Y_i) form a Poisson random measure, Q_t is the number of them in the wedge A_t.

This defines a random measure L on $(E, \mathcal{E})$; it is said to be a compound Poisson random measure. This is a slight generalization of the processes with the similar name.

a) Show that $L(A_1), \ldots, L(A_n)$ are independent whenever $A_1, \ldots, A_n$ are disjoint sets in $\mathcal{E}$.

b) Compute the mean and the Laplace functional of L.

3.35 *Infinite server queues.* Consider a service facility with an infinite number of servers, a Poisson process of customer arrivals with intensity a, and independent service times with a common distribution π. Let X_i be the arrival time of the customer labeled i, let Y_i be the corresponding service time. It is assumed that X and Y satisfy the conditions of Corollary 3.5 with $E = \mathbb{R}$, $F = \mathbb{R}_+$, $\nu = a$ Leb, and π as the service distribution here.

a) Show that the departure times $X_i + Y_i$ form a Poisson random measure on $\mathbb{R}$ with mean $\nu = a$ Leb.

b) Let Q_t be the number of customers in the system at time t; this is the number of pairs (X_i, Y_i) in the wedge $A_t = \{(x, y) \in \mathbb{R} \times \mathbb{R}_+ : x \leq t < x + y\}$. What is the distribution of Q_t? Describe the joint distribution of Q_s and Q_t for times $s < t$; what is the covariance. See Figure 8 above.

c) Suppose that the customer i pays an amount $f(Y_i)$ at his departure time $X_i + Y_i$. Consider the present worth W of all payments after time 0 if the discount rate is r, that is,

$$W = \sum_i f(Y_i)\ e^{-r(X_i+Y_i)}\ 1_{\mathbb{R}_+}(X_i + Y_i).$$

Compute the mean, variance, and Laplace transform of W.

d) Re-do the answers assuming that the arrival process has intensity $a(x)$ at time x and the service distribution is $\pi(x, \cdot)$ depending on the arrival time x.

3.36 *Continuation.* Think of the service facility as a large national park. Customer i arrives at time X_i and stays in the park for Y_i units of time and, during the interval $[X_i, X_i + Y_i)$ the customer moves in the forest F according to some process $Z_i = \{Z_i(u) : 0 \le u \le Y_i\}$. Let $Q_t(B)$ be the number of customers in the subset B of F at time t. Characterize Q_t under reasonable assumptions regarding the motions Z_i.

3.37 *Traffic flow.* This is a reasonable account of low density traffic flows. It applies well in situations where vehicles can pass each other freely, and, therefore, the speed of a vehicle remains nearly constant over time and does not depend on the speeds or positions of other vehicles. Under these conditions, it can be shown that the configuration of vehicles on the road approximates a Poisson configuration.

We take the road to be $\mathbb{R}$. Vehicles are labeled with the integers, X_i is the position at time 0 of the vehicle i, and V_i is its velocity (random, but constant over time). We suppose that $X = \{X_i : i \in \mathbb{Z}\}$ forms a Poisson random measure on $\mathbb{R}$ with mean $\nu = c$ Leb, and that $V = \{V_i : i \in \mathbb{Z}\}$ is independent of X and is an independency of variables with the same distribution π on $\mathbb{R}$.

a) Describe the random measure N_t representing the configuration of vehicles at time t, that is, N_t is formed by $X_i + tV_i$, $i \in \mathbb{Z}$.

b) Fix a point x on the highway. Let M_x be formed by the times T_i at which vehicles pass by the point x. Show that it is Poisson on $\mathbb{R}$ with intensity $a = cb$ where $b = \mathbb{E}\,|V_i|$.

c) Consider the relative motions of the vehicles with respect to a marked (observer's) vehicle. Suppose that the marked vehicle is at x_0 at time 0 and travels with the constant deterministic velocity v_0 in $(0, \infty)$. Assume that all the V_i are positive (that is, we are considering the traffic moving in the positive direction). Let M_a be formed by the times at which vehicles with speeds above v_0 pass the marked one, and M_b by the times at which the marked vehicle passes the vehicles with speeds below v_0. Show that M_a and M_b are independent Poisson random measures on $\mathbb{R}_+$. Find their mean measures. Safety hint: What should v_0 be in order to minimize the expected rate of passings of both kinds?

3.38 *Continuation.* We now put in entrances and exits on the highway. The vehicle i enters the highway at time T_i at the point X_i, moves with speed V_i, and exits at point Y_i. It is reasonable to assume that the T_i form a Poisson random measure with mean $\mu(dt)$, the pairs (X_i, Y_i) have a joint distribution $Q(t, dx, dy)$ depending on time $t = T_i$, and that the V_i has the distribution $R(t, x, y, dv)$ depending on $T_i = t$, $X_i = x$, $Y_i = y$. Make these precise. Do the same problems as in 3.37.

3.39 *Stereology.* A piece of metal contains chunks of some foreign substance distributed in it at random. A plane section of the metal is inspected and the disks of foreign substance seen are counted and their radii are noted. The problem is to infer, from such data, the volume occupied by the foreign

substance. A similar problem arises in studies of growth of cancerous tissue. Mice are injected with carcinogens when they are a few days old, and their livers are taken out for inspection a few weeks after. Each liver is sliced, and sizes and locations of cancerous parts (seen on that plane) are measured. The problem is to infer the volume of liver occupied by the cancerous tissue. This exercise is about the essential issues.

Let X and Y be as in Corollary 3.5, but with $E = \mathbb{R}^3$, $F = \mathbb{R}_+$, $\nu = c\,\text{Leb}$, and π a distribution on $\mathbb{R}_+$. We replace Y with R for our present purposes. According to the corollary, then, (X, R) forms a Poisson random measure M on $\mathbb{R}^3 \times \mathbb{R}_+$ with mean $\mu = c\,\text{Leb}\times\pi$. We think of an atom (X_i, R_i) as a ball of radius R_i centered at X_i.

Consider the intersections of these balls with the plane $\{(x, y, z) \in \mathbb{R}^3 : z = 0\}$. A ball with center (x, y, z) and radius r intersects this plane if and only if $r > |z|$, and if it does, then the intersection is a disc of center (x, y) and radius $q = \sqrt{r^2 - z^2}$. Let L be the random measure on $\mathbb{R}^2 \times \mathbb{R}_+$ formed by such "disks". Show that L is Poisson with mean measure $bc\,\text{Leb}\times\lambda$, where b is the mean radius of a ball and λ is the probability measure (for the disk radius)

$$\lambda(dq) = \frac{1}{b}\text{Leb}(dq) \int_q^\infty \pi(dr) \frac{q}{\sqrt{r^2 - q^2}}.$$

Let V_t be the total volume of all the balls whose centers are within a distance t from the origin, and let A_t be the total area of all the disks on the plane $z = 0$ whose centers are within a distance t from the origin. Compute the ratio $(\mathbb{E}\, V_t)/(\mathbb{E}\, A_t)$.

3.40 *Poisson fields of lines.* Consider a system of random lines in $\mathbb{R}^2$. We are interested in the sizes and shapes of polygons formed by those lines. In Exercise 3.37 on traffic flow, if we draw the paths of all the vehicles on a time $\times$ space coordinate system, the resulting collection of lines would be an example. Since each line corresponds to an atom (X_i, V_i) of a Poisson random measure, it is appropriate to call the system a Poisson field of lines. In metallurgy, in discussions of sizes and shapes of grains that make up the granular structure of the metal, the grains are approximated by polygons formed by such lines (the model is poor, but it seems to be the only tractable one).

Let g be an (infinite straight) line in $\mathbb{R}^2$. By its *distance* from the origin we mean the length of the perpendicular drawn from the origin to the line. By its *orientation* is meant the angle that the perpendicular makes with the x-axis. Orientation is an angle between 0 and π, distance is positive or negative depending on whether the perpendicular is above or below the x-axis. Note that, if g has distance d and orientation α then

$$(x, y) \in g \iff d = x\cos\alpha + y\sin\alpha.$$

Let (D_i) form a Poisson random measure N on $\mathbb{R}$ with mean measure $\nu = \pi c\,\text{Leb}$ where $c > 0$ is fixed; let (A_i) be independent of (D_i), and the A_i be independent of each other with uniform distribution on $[0, \pi]$.

a) Describe the random measure M formed by the pairs (D_i, A_i) on $\mathbb{R} \times [0, \pi]$. The lines G_i corresponding to the pairs (D_i, A_i) are said to form a Poisson field of lines with intensity c.

b) Let X_i be the intersection of the line G_i with the x-axis and B_i the angle (between 0 and π) between G_i and the x-axis. Show that (X_i, B_i) form a Poisson random measure on $\mathbb{R} \times [0, \pi]$ with mean measure $2c\ dx\ \beta(db)$, where β is the distribution $\beta(db) = \frac{1}{2}(1 - \cos b)$, $0 \le b \le \pi$.

c) The Poisson field of lines G_i is invariant in law under translations and rotations of the plane. Show this by considering the translation $(x, y) \to (x_0 + x, y_0 + y)$ for some fixed $(x_0, y_0) \in \mathbb{R}^2$, and then by considering the rotation of the plane by an angle α_0.

d) The number K of random lines intersecting a fixed convex region $C \subset \mathbb{R}^2$ with perimeter p has the Poisson distribution with mean cp. Show this. Hints: Whether a line G_i intersects C is merely a function of (D_i, A_i). If G_i intersects C, it intersects C twice. So $2K$ is the total number of intersections. Suppose C is a polygon with n sides of lengths $p_1, \ldots, p_n$ (then the perimeter is $p = p_1 + \cdots + p_n$). Use the results of (b) and (c) above to compute the expected numbers of intersections with each side. The sum of these numbers is $2\mathbb{E}\ K$. Finally, approximate C by polygons.

e) This requires much knowledge of geometry. The lines G_i partition the plane into random polygons (some of these are triangles, some have four sides, some 17, etc.). Take an arbitrary one of these (can you make this precise so that the following are random variables?). Let S be the number of sides, P the perimeter, A the area, D the diameter of the in-circle. Then, D has the exponential distribution with parameter πc, and $\mathbb{E}\ S = 4$, $\mathbb{E}\ P = 2/c$, $\mathbb{E}\ A = 1/\pi c^2$.

4 Additive Random Measures and Lévy Processes

Our aim is to illustrate the uses of Poisson random measures to construct more evolved random measures. These are related intimately to Lévy processes to be studied in the next chapter; we give a few examples here.

Throughout, $(E, \mathcal{E})$ will be a fixed measurable space. If μ is a measure on a measurable space $(F, \mathcal{F})$, we shall omit mentioning the σ-algebra and merely say that μ is a measure on F. This convention extends to random measures naturally. Indeed, on spaces such as $\mathbb{R}_+$ or $\mathbb{R}^d$, we shall omit specifying the σ-algebra; they will always be the Borel σ-algebras.

4.1 Definition. *Let M be a random measure on E. It is said to be* additive *if $M(A), \ldots, M(B)$ are independent for all choices of the finitely many disjoint sets $A, \ldots, B$ in $\mathcal{E}$.*

Every deterministic measure is additive. Every Poisson random measure is additive. We shall see shortly that archetypical additive random measures on E are constructed from Poisson random measures on $E \times \mathbb{R}_+$.

The probability law of an additive random measure M is specified once the distribution of $M(A)$ is specified for each A in $\mathcal{E}$. To see this, note that the joint distribution of $M(B)$ and $M(C)$ can be computed from the marginal distributions of $M(B \setminus C)$, $M(B \cap C)$, $M(C \setminus B)$ using the independence of the last three variables, and that this argument extends to finite-dimensional distributions of M.

Construction of additive measures

The following seems to require no proof. It shows the construction of a purely atomic additive random measure whose atoms are fixed; only the weights on the atoms are random.

4.2 Lemma. *Let D be a countable subset of E, and let $\{W_x : x \in D\}$ be an independency of positive random variables. Define*

$$K(\omega, A) = \sum_{x \in D} W_x(\omega)\ I(x, A), \quad \omega \in \Omega, \quad A \in \mathcal{E}.$$

Then, K is an additive random measure.

At the other extreme, a Poisson counting measure has weight one on each of its atoms, but the atoms themselves are generally random. The following is the construction of additive measures of general interest.

4.3 Lemma. *Let N be a Poisson random measure on $E \times \mathbb{R}_+$ with mean measure ν. Define*

$$L(\omega, A) = \int_{A \times \mathbb{R}_+} N(\omega; dx, dz)\ z, \quad \omega \in \Omega, \quad A \in \mathcal{E}.$$

Then, L is an additive random measure on E. The Laplace transform for $L(A)$ is, for A in $\mathcal{E}$,

$$\mathbb{E}\ e^{-rL(A)} = \exp_- \int_{A \times \mathbb{R}_+} \nu(dx, dz)\ (1 - e^{-rz}), \quad r \in \mathbb{R}_+.$$

Proof. By Fubini's theorem, L is a random measure. Note that $L(A)$ is determined by the trace of N on $A \times \mathbb{R}_+$. If $A, \dots, B$ are finitely many disjoint sets in $\mathcal{E}$, then the traces of N over $A \times \mathbb{R}_+, \dots, B \times \mathbb{R}_+$ are independent by Exercise 2.31, and, hence, $L(A), \dots, L(B)$ are independent. So, L is additive. The formula for the Laplace transform follows from Theorem 2.9 on the Laplace functional of N after noting that $rL(A) = Nf$ with $f(x, z) = rz1_A(x)$. □

4.4 Theorem. *Let α be a deterministic measure on E. Let K be as in Lemma 4.2 and L as in Lemma 4.3, and suppose that K and L are independent. Then,*

$$M = \alpha + K + L$$

is an additive random measure on E.

Proof. It is immediate from the lemmas above and the observation that the sum of independent additive random measures is again additive. □

Conversely, it can be shown that the preceding is, basically, the general form of an additive random measure. More precisely, if M is an additive random measure on a standard measurable space $(E, \mathcal{E})$ and it is Σ-bounded as a kernel, then it has the decomposition $M = \alpha + K + L$ with the components as described in the preceding theorem; moreover, this decomposition is unique provided that both α and $\nu(\cdot \times \mathbb{R}_+)$ be diffuse measures on E, that is, all the fixed atoms (if any) belong to K.

Increasing Lévy processes

This is to establish a connection between additive random measures and the processes to be studied in the next chapter in greater generality. We start by introducing an important class of such processes.

4.5 DEFINITION. *Let $S = (S_t)_{t\in\mathbb{R}_+}$ be an increasing right-continuous stochastic process with state space $\mathbb{R}_+$ and $S_0 = 0$. It is said to be an* increasing Lévy process *(or* subordinator*) if*

a) the increments $S_{t_1} - S_{t_0}, S_{t_2} - S_{t_1}, \ldots, S_{t_n} - S_{t_{n-1}}$ are independent for $n \geq 2$ and $0 \leq t_0 < t_1 < \cdots < t_n$, and

b) the distribution of the increment $S_{t+u} - S_t$ is the same as that of S_u for every t and u in $\mathbb{R}_+$.

The property (a) is called the independence of increments, and (b) the stationarity of the increments. More general Lévy processes are defined by the properties (a) and (b), but for right-continuous left-limited processes with state spaces $\mathbb{R}^d$; see Chapter VII.

Given an additive random measure M on $\mathbb{R}_+$, putting $S_t(\omega) = M(\omega, [0, t])$ yields an increasing right-continuous process. Once we assure that $S_t < \infty$ almost surely for all t, independence of increments follows from the additivity of M. Stationarity of increments is achieved by making sure that the mean measure is chosen appropriately and there be no fixed atoms and the deterministic measure α be a constant multiple of the Lebesgue measure. In other words, the following proposition is in fact a complete characterization of increasing Lévy processes; here we state and prove the sufficiency part, the necessity will be shown in the next chapter; see also 4.13 below.

4.6 PROPOSITION. *Let b be a constant in $\mathbb{R}_+$. Let N be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ whose mean has the form $\nu = Leb \times \lambda$, where the measure λ satisfies*

4.7
$$\int_{\mathbb{R}_+} \lambda(dz)\ (z \wedge 1)\ < \infty.$$

Define

$$S_t(\omega) = bt + \int_{[0,t]\times\mathbb{R}_+} N(\omega; dx, dz)\ z, \quad t \in \mathbb{R}_+, \quad \omega \in \Omega.$$

Then, $S = (S_t)_{t\in\mathbb{R}_+}$ *is an increasing Lévy process, and*

4.8 $$\mathbb{E}\ e^{-rS_t} = \exp_-\ t\left[br + \int_{\mathbb{R}_+} \lambda(dz)(1-e^{-rz})\right], \quad r \in \mathbb{R}_+.$$

REMARK. Let L be defined on $E = \mathbb{R}_+$ as in Lemma 4.3 from the Poisson random measure N here. Let $\alpha = b$ Leb on $E = \mathbb{R}_+$. Then, $M = \alpha + L$ is an additive random measure on $\mathbb{R}_+$, and we have $S_t(\omega) = M(\omega, [0,t])$ for all t and ω.

Proof. It is obvious that S is increasing, right-continuous, and $S_0 = 0$. Note that $S_t = bt + Nf$, where $f(x,z) = z1_{[0,t]}(x)$. In view of 4.7,

$$\nu(f\wedge 1) = \int_{\mathbb{R}_+\times\mathbb{R}_+} \nu(dx,dz)(f(x,z)\wedge 1) = t\int_{\mathbb{R}_+} \lambda(dz)(z\wedge 1)$$

is finite, which shows that Nf and, thus, S_t are almost surely finite. Now, the increments are well-defined, and their independence follows from the Poisson nature of N. The formula for the Laplace transform is immediate from Theorem 2.9 with the present ν and f. Finally, a similar computation shows that the Laplace transforms of $S_{t+u} - S_t$ and S_u are the same, and hence the stationarity of increments. □

The constant b is called the drift coefficient, and the measure λ the *Lévy measure*, of the process S. Obviously, together, they determine the probability law of S. We shall give two particular examples of such λ below. For the present, we add that every finite measure λ on $\mathbb{R}_+$ satisfies the condition 4.7 for the Lévy measure.

Examples

4.9 *Gamma processes.* Let S be as in the last proposition with $b = 0$ and

$$\lambda(dz) = dz \cdot a\frac{e^{-cz}}{z}, \quad z > 0$$

for some constants a and c in $(0,\infty)$; note that 4.7 is satisfied. Thus, S is an increasing Lévy process. Now, 4.8 becomes

$$\begin{aligned}\mathbb{E}\ e^{-rS_t} &= \exp_-\ t\int_0^\infty dz\ a\ \frac{e^{-cz}}{z}(1-e^{-rz}) \\ &= \exp_-\ t\ a\log\frac{c+r}{c} = \left(\frac{c}{c+r}\right)^{at};\end{aligned}$$

thus, S_t has the gamma distribution with shape index at and scale parameter c. For this reason, S is said to be a *gamma process with shape rate* a and scale parameter c.

4.10 *Increasing stable processes.* Let S be as in Proposition 4.6 again, but with $b = 0$ and

$$\lambda(dz) = dz\ \frac{ac}{\Gamma(1-a)}\ z^{-1-a}, \quad z > 0,$$

where a is a constant in $(0,1)$, and c is a constant in $(0,\infty)$, and Γ denotes the gamma function. Again, λ satisfies the condition 4.7, and S is an increasing Lévy process. Even though $S_t < \infty$ almost surely for every t in $\mathbb{R}_+$,

$$\mathbb{E}\ S_t = t \int_{\mathbb{R}_+} \lambda(dz) z = t \cdot (+\infty) = +\infty$$

for every $t > 0$, however small t may be. This process S is said to be *stable with index* a *in* $(0,1)$. *Stability* refers to the fact that $(S_{ut})_{t\in\mathbb{R}_+}$ has the same probability law as $(u^{1/a} S_t)_{t\in\mathbb{R}_+}$ for every $u > 0$; this can be seen by recalling that the probability law of an additive measure is specified by its one-dimensional distributions and that

$$\mathbb{E}\ e^{-rS_t} = e^{-tcr^a}, \quad t \in \mathbb{R}_+, \quad r \in \mathbb{R}_+,$$

in view of the formula 4.8 and the form of λ here.

The distribution of S_t does not have an explicit form in general. However, for $a = 1/2$, we have

$$\mathbb{P}\{S_t \in dz\} = dz \cdot \frac{ct}{\sqrt{4\pi z^3}}\ e^{-c^2t^2/4z}, \quad z > 0.$$

The further special case where $a = 1/2$ and $c = \sqrt{2}$ plays an important role in the theory of Brownian motion: as can be seen in Proposition V.5.20, there, S_x becomes the time it takes for the Wiener process to go from 0 to x.

4.11 *Stable random measures.* These are generalizations of the preceding example. Let L be as in Lemma 4.3, but with $E = \mathbb{R}^d$ for some dimension $d \geq 1$. Assume that the mean of the Poisson random measure there has the form $\nu = \text{Leb} \times \lambda$, where the measure λ on $\mathbb{R}_+$ is as in Example 4.10 above. The formula for the Laplace transform in 4.3 becomes

$$\mathbb{E}\ e^{-rL(A)} = \exp_(\text{Leb}\ A) cr^a, \quad r \in \mathbb{R}_+.$$

If Leb $A = \infty$, then $L(A) = +\infty$ almost surely. If Leb $A < \infty$, then $L(A) < \infty$ almost surely. But $\mathbb{E}\ L(A) = +\infty$ for every A with Leb $A > 0$, however small its measure may be. This additive random measure on $\mathbb{R}^d$ is *stable* in the following sense: Let u be a constant in $(0,\infty)$ and note that the Lebesgue measure of $uA = \{ux : x \in A\}$ is equal to u^dLeb A. Thus, the Laplace transform of $\hat{L}(A) = u^{-d/a} L(uA)$ is the same as that of $L(A)$. This means

that the probability laws of the additive random measures L and $\hat{L}$ are the same. In other words, the probability law of L remains *stable* under the transformation $(x, z) \mapsto (ux, u^{-d/a}z)$ of $\mathbb{R}^d \times \mathbb{R}_+$ into itself, this being true for every scalar $u > 0$.

4.12 *Gamma random measures.* These are counterparts, on arbitrary spaces, of the processes of Example 4.9. Let L and N be as in Lemma 4.3 with $(E, \mathcal{E})$ arbitrary, but the mean ν of N having the form $\nu = \mu \times \lambda$, with μ an arbitrary measure on E, and λ the measure on $\mathbb{R}_+$ given in Example 4.9. Then, L is an additive random measure on E, and

$$\mathbb{E}\, e^{-rL(A)} = \exp_-\ \mu(A) \int_{\mathbb{R}_+} \lambda(dz)(1 - e^{-rz}) = \left(\frac{c}{c+r}\right)^{a\mu(A)}.$$

If $\mu(A) = +\infty$ then $L(A) = \infty$ almost surely. If $\mu(A) < \infty$, then $L(A)$ is almost surely finite and has the gamma distribution with shape index $a\mu(A)$ and scale c. Example 4.9 is, basically, the special case where $E = \mathbb{R}_+$ and $\mu = \text{Leb}$.

Homogeneity and stationarity

Suppose that E is $\mathbb{R}_+$ or $\mathbb{R}^d$, and $\mathcal{E}$ is the corresponding Borel σ-algebra. An additive random measure on E is said to be *homogeneous* if its probability law remains invariant under shifts of the origin in E. If E is $\mathbb{R}_+$ or $\mathbb{R}$ and is regarded as time, the term *stationary* is preferred instead.

Let M be an additive random measure on E. Then, its probability law is determined by specifying the distribution of $M(A)$ for each A is $\mathcal{E}$. Thus, homogeneity of M is equivalent to requiring that, for every A in $\mathcal{E}$, the distribution of $M(x+A)$ remains the same while x ranges over E; here and below, $x + A = \{x + y : y \in A\}$. This observation is the primary ingredient in the proof of the following proposition, which is a slightly more general version of random measures associated with increasing Lévy processes of Proposition 4.6 above. Recall that the kernel M is σ-bounded if there is a countable partition (A_n) of E such that $M(A_n) < \infty$ almost surely for each n.

4.13 PROPOSITION. *Suppose that E is $\mathbb{R}_+$ or $\mathbb{R}^d$. Let M be a σ-bounded additive random measure on E as in Theorem 4.4. Then, M is homogeneous if and only if it has the form*

$$M(\omega, A) = b\ \mathit{Leb}\ A + \int_{A\times\mathbb{R}_+} N(\omega; dx, dz)z, \quad \omega \in \Omega, \quad A \in \mathcal{E},$$

for some constant b in $\mathbb{R}_+$ and some Poisson random measure N on $E \times \mathbb{R}_+$ with mean $\nu = \mathit{Leb} \times \lambda$, where λ satisfies 4.7.

Proof. Sufficiency. Let M be as described. It is obviously additive. To show that M is σ-bounded, it is enough to show that $M(A) < \infty$ almost surely for every A in $\mathcal{E}$ with Leb $A < \infty$; then, the σ-finiteness of the Lebesgue

measure does the rest. So, let A be such that $c = \text{Leb}\, A$ is finite. Note that $M(A) = bc + Nf$ where $f(x,z) = z1_A(x)$, and that

$$\nu(f \wedge 1) = c \int_{\mathbb{R}_+} \lambda(dz)(z \wedge 1) < \infty$$

by condition 4.7 on λ. Thus, $Nf < \infty$ almost surely by Proposition 2.13, and thus $M(A) < \infty$ almost surely. Finally, M is homogeneous since $M(A)$ and $M(x + A)$ have the same distribution: we have

$$\mathbb{E}\, e^{-rM(A)} = \exp_ \ (\text{Leb}\, A)\left(br + \int_{\mathbb{R}_+} \lambda(dz)(1 - e^{-rz})\right),$$

and $\text{Leb}\,(x + A) = \text{Leb}\, A$.

Necessity. Suppose that $M = \alpha + K + L$ as in Theorem 4.4 and, in addition, is homogeneous and σ-bounded. The homogeneity has two consequences:

First, M cannot have fixed atoms. If there were a fixed atom x_0 then $x + x_0$ would be a fixed atom for every x in E, which contradicts the σ-boundedness of M (recall that a σ-finite measure can have at most countably many atoms). Thus, $K = 0$, and α is diffuse, and the Poisson random measure N defining L has no fixed atoms.

Second, homogeneity implies that $M(A)$ and $M(x + A)$ have the same distribution for all x and A. Thus, $\alpha(A) = \alpha(x + A)$ and $\nu(A \times B) = \nu((x + A) \times B)$ for all x, A, B appropriate. These imply that $\alpha = b\,\text{Leb}$ for some constant b in $\mathbb{R}_+$ and that $\nu = \text{Leb} \times \lambda$ for some measure λ on $\mathbb{R}_+$.

Finally, σ-boundedness of M implies that there is A in $\mathcal{E}$ with $c = \text{Leb}\, A$ belonging to $(0, \infty)$ such that $M(A) < \infty$ almost surely. Since $M(A) = bc + Nf$ with $f(x,z) = z1_A(x)$, we must have $Nf < \infty$ almost surely, which in turn implies via Proposition 2.13 that

$$\nu(f \wedge 1) = c \int_{\mathbb{R}_+} \lambda(dz)(z \wedge 1)$$

must be finite. Hence, the condition 4.7 holds on λ. □

Alternative constructions

Consider an increasing Lévy process S constructed as in Proposition 4.6. The condition 4.7 on λ implies that $\lambda(\varepsilon, \infty) < \infty$ for every $\varepsilon > 0$, and it follows from Proposition 2.18 that no two atoms of N arrive at the same time, that is, for almost every ω, if (t, z) and (t', z') are distinct atoms of the counting measure $N(\omega, \cdot)$, then t and t' are distinct as well. If follows that, for almost every ω, the path $t \mapsto S_t(\omega)$ has a jump of size z at time t if and only if (t, z) is an atom of $N(\omega, \cdot)$.

Fix $z > 0$ and consider the times at which S has a jump of size exceeding z. Those times form a Poisson random measure on $\mathbb{R}_+$ whose mean measure

is $\lambda(z,\infty)\cdot$ Leb; in other words, $\lambda(z,\infty)$ is the time rate of jumps of size exceeding z. More generally, $\lambda(B)$ is the rate of jumps of size belonging to B.

In the construction of Proposition 4.6, then, the burden of describing the probabilistic jump structure is on the measure λ, which in turn specifies the probability law of N. Next, we describe a construction where the law of N is parameter free and the jump mechanism is more explicit: to every atom (t,u) of $N(\omega,\cdot)$ there corresponds a jump of size $j(u)$ at time t. We ignore the drift term.

4.14 PROPOSITION. *Let N be a standard Poisson random measure on $\mathbb{R}_+\times\mathbb{R}_+$ (with mean Leb $\times$ Leb). Let $j:\mathbb{R}_+\mapsto\mathbb{R}_+$ be a Borel function satisfying*

4.15 $$\int_{\mathbb{R}_+} du\ (j(u)\wedge 1)<\infty.$$

Define

$$S_t(\omega)=\int_{[0,t]\times\mathbb{R}_+} N(\omega;dx,du)\ j(u),\quad \omega\in\Omega,\quad t\in\mathbb{R}_+.$$

Then, $S=(S_t)$ is an increasing Lévy process with

4.16 $$\mathbb{E}\ e^{-rS_t}=\exp_\ t\int_0^\infty du\ (1-e^{-rj(u)}),\quad r\in\mathbb{R}_+.$$

Proof. Observe that $S_t=Nf$ where $f(x,u)=j(u)\,1_{[0,t]}(x)$. The condition 4.15 ensures that $Nf<\infty$ almost surely. That S is an increasing Lévy process follows from the Poisson character of N and the particular mean measure. Finally, the formula 4.16 is immediate from Theorem 2.9 on the Laplace functionals. □

Let j be as in the preceding proposition and define

4.17 $$\lambda=\text{Leb}\circ j^{-1}.$$

Then, the condition 4.15 on the jump function j is equivalent to the condition 4.7 on λ, the Laplace transforms 4.16 and 4.8 are the same, and, hence, the processes S of Propositions 4.14 and 4.6 have the same probability law. These conclusions remain true when, in the converse direction, we start with the measure λ and define j so that 4.17 holds. This is easy to do at least in principle: since λ satisfies 4.7, the function $z\mapsto\lambda(z,\infty)$ is real-valued and decreasing over $(0,\infty)$, and defining

4.18 $$j(u)=\inf\{z>0:\lambda(z,\infty)\le u\},\quad u\in\mathbb{R}_+,$$

we see that 4.17 holds.

4.19 Example. *Stable processes.* Let λ be as in Example 4.10 and recall that $a \in (0,1)$. Then, $\lambda(z,\infty) = c/z^a\Gamma(1-a)$, and 4.18 yields $j(u) = \hat{c}u^{-1/a}$ for all $u > 0$ with $\hat{c} = (c/\Gamma(1-a))^{1/a}$. With this j, the process S of the last proposition has the same probability law as that of Example 4.10. Incidentally, in the particular case where $a = 1/2$ and $c = \sqrt{2}$, we have

$$\lambda(dz) = dz \frac{1}{\sqrt{2\pi z^3}}, \quad j(u) = \frac{2}{\pi u^2}. \qquad \square$$

In Monte Carlo studies, the construction of Proposition 4.14 is preferred over that of Proposition 4.6, because constructing a standard Poisson random measure is easier; see Remark 2.16. More importantly, 4.14 enables us to construct two (or any number of) processes using the same standard Poisson random measure, but with different jump functions for different processes; this is advantageous in making stochastic comparisons.

The practical limitations to the method of the last proposition come from the difficulties in implementing 4.18. For instance, this is the case for the gamma process of Example 4.9 because the tail $\lambda(z,\infty)$ does not have an explicit expression in that case. Every such instance requires special numerical methods. The following is an analytical solution in the case of gamma processes.

4.20 Example. *Gamma processes.* Let $\lambda(dz) = dz\ ae^{-cz}/z$, $z > 0$, the same as in Example 4.9. The formula 4.18 for j is difficult to use because $\lambda(z,\infty)$ does not have an explicit expression. However, it is possible to write

4.21 $$\lambda = \mu \circ j^{-1}$$

for a pleasant measure μ and explicit j, but μ and j must be defined on $\mathbb{R}_+ \times \mathbb{R}_+$: for (u,v) in $\mathbb{R}_+ \times \mathbb{R}_+$, let

4.22 $$\mu(du,dv) = du\ dv\ ace^{-cv}, \quad j(u,v) = e^{-u}v.$$

It is easy to check that 4.21 holds. The measure μ can be thought as $\mu = (a\ \text{Leb}) \times \eta$ where η is the exponential distribution on $\mathbb{R}_+$ with parameter c. Thus, constructing a Poisson random measure N on $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$ with mean $\text{Leb} \times \mu$ is easy, and the following construction should be preferred over that in Example 4.9. For motivation and comments on this construction we refer to Exercises 4.29 and 4.30 below.

Let N be a Poisson random measure on $\mathbb{R}_+ \times (\mathbb{R}_+ \times \mathbb{R}_+)$ with mean $\nu = \text{Leb} \times \mu$. Recalling $j(u,v) = e^{-u}v$, define

$$S_t(\omega) = \int_{[0,t]\times\mathbb{R}_+\times\mathbb{R}_+} N(\omega; dx, du, dv)\ j(u,v), \quad t \in \mathbb{R}_+, \quad \omega \in \Omega.$$

Then, $S = (S_t)$ is a gamma process with shape rate a and scale parameter c, just as the process S in Example 4.9.

To see the truth of the last assertion, it is enough to note that S is an increasing Lévy process obviously, and then note that the Laplace transform of S_t is

$$\mathbb{E}\, e^{-rS_t} = \exp_- \int_{[0,t]} dx \int_{\mathbb{R}_+} du \int_{\mathbb{R}_+} dv\ ace^{-cv}\ (1 - e^{-re^{-u}v}) = \left(\frac{c}{c+r}\right)^{at}$$

as needed. Another, simpler, way of seeing the result is by noting that

$$S_t = \int_{[0,t]\times\mathbb{R}_+} \hat{N}(dx, dz)\ z,$$

where $\hat{N} = N \circ h^{-1}$ with $h(x, u, v) = (x, e^{-u}v) = (x, j(u, v))$. Then, $\hat{N}$ is Poisson with mean $\nu \circ h^{-1} = (\text{Leb} \times \mu) \circ h^{-1} = \text{Leb} \times \lambda$ since $\mu \circ j^{-1} = \lambda$, which shows that $\hat{N}$ has the same law as the Poisson random measure in Example 4.9. □

Exercises

4.23 *Additive measures with fixed atoms.* Let K be as in Lemma 4.2, but with $E = [0, 1]$ and D the set of all strictly positive rational numbers in E. Suppose that the independent variables W_x, $x \in D$, are exponentially distributed with means m_x, $x \in D$, chosen such that $\sum_{x\in D} m_x = 1$. Then K is an additive random measure whose atoms are fixed. Show that K is almost surely finite. Compute the Laplace functional $\mathbb{E}\, e^{-Kf}$ for positive Borel functions f on E.

4.24 *Continuation.* Choose the numbers m_x for rational numbers x in $(0, 1]$ such that their sum is equal to 1. For rational x in $(n, n + 1]$, define $m_x = m_{x-n}$. Let W_x have the exponential distribution with mean m_x for every x in the set D of all strictly positive rationals, and assume that the W_x are independent. Let K be defined as in Lemma 4.2 with $E = \mathbb{R}_+$ and this D. Then, K is an additive random measure on $\mathbb{R}_+$. Show that it is σ-bounded. Define $S_t(\omega) = K(\omega, [0, t])$ for ω in Ω and t in $\mathbb{R}_+$. Show that $S = (S_t)$ is a strictly increasing right-continuous process with state space $\mathbb{R}_+$. Show that it has independent increments, that is, the condition (a) of Definition 4.5 holds, but the increments are not stationary, that is, the condition (b) does not hold.

4.25 *Continuation.* Let S be as in the preceding exercise. Let f be an arbitrary Borel function on $\mathbb{R}_+$. Define $\hat{S}_t(\omega) = f(t) + S_t(\omega)$. Show that $\hat{S}$ has independent increments. If f fails to be right-continuous, $\hat{S}$ will fail to be right-continuous.

4.26 *Laplace functionals.* Let L be as in Lemma 4.3. Show that, for every f in $\mathcal{E}_+$,

$$\mathbb{E}\, e^{-Lf} = \exp_- \int_{E\times\mathbb{R}_+} \nu(dx, dz)\ (1 - e^{-zf(x)}).$$

4.27 *Compound Poisson processes.* Show that the condition 4.7 is satisfied by every finite measure λ on $\mathbb{R}_+$. Let S be defined as in Proposition 4.6 but with $b = 0$ and λ finite with $\lambda\{0\} = 0$.

a) Show that the atoms of N can be labeled as $(T_1, Z_1), (T_2, Z_2), \ldots$ so that $0 < T_1 < T_2 < \cdots$ almost surely. Show that the T_i form a Poisson random measure on $\mathbb{R}_+$ with mean equal to c Leb, where $c = \lambda(0, \infty)$; see 1.10 *et seq.*

b) Show that the paths $t \mapsto S_t(\omega)$ are increasing step functions; the jumps occur at $T_1(\omega), T_2(\omega), \ldots$; the size of the jump at $T_i(\omega)$ is equal to $Z_i(\omega)$; between the jumps the paths remain constant.

4.28 *Compound Poisson continued.* Let N be a standard Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ (with mean Leb$\times$Leb). Let $j : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be a Borel function with a compact support, that is, there is b in $(0, \infty)$ such that j vanishes outside $[0, b]$. Define

$$S_t = \int_{[0,t]\times\mathbb{R}_+} N(dx, du)\; j(u), \quad t \in \mathbb{R}_+.$$

Show that S is a compound Poisson process.

4.29 *Gamma distribution.* Let (U_i) form a Poisson random measure on $\mathbb{R}_+$ with mean a Leb; see 1.10 *et seq.* for the terms. Let (V_i) be independent of (U_i) and be an independency of exponential variables with parameter c. Show that

$$X = \sum_i e^{-U_i} V_i$$

has the gamma distribution with shape index a and scale parameter c. Hint: Let N be the Poisson random measure formed by the pairs (U_i, V_i), and note that $X = Nf$ with $f(u, v) = e^{-u}v$ to compute the Laplace transform of X.

4.30 *Gamma processes.* Let the pairs (T_i, U_i) form a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean equal to a Leb $\times$ Leb. Let the family (V_i) be an independency of exponential variables with parameter c. Assume that the collections (V_i) and (T_i, U_i) are independent. Define

$$S_t = \sum_i e^{-U_i} V_i \; 1_{[0,t]} \circ T_i, \quad t \in \mathbb{R}_+.$$

Show that (S_t) is a gamma process with shape rate a and scale parameter c. This representation is equivalent to that of Example 4.20.

4.31 *Lévy measure as an image.* Let λ, μ, and j be as in Example 4.20. Show the claim that $\lambda = \mu \circ j^{-1}$. It is possible to change μ and j while keeping intact the relationship $\lambda = \mu \circ j^{-1}$. Show this for

a) $\mu(du, dv) = du\; dv\; e^{-v}$ and $j(u, v) = e^{-u/a} v/c$, $(u, v) \in \mathbb{R}_+ \times \mathbb{R}_+$,

b) $\mu(du, dv) = du\; dv$ and $j(u, v) = e^{-u/a}(-\log v)/c$, $(u, v) \in \mathbb{R}_+ \times (0, 1)$.

4.32 *Another construction.* This is the generalization of the foregoing ideas to arbitrary additive random measures. Let N be a Poisson random measure on some measurable space $(D, \mathcal{D})$. Let $j : D \mapsto \mathbb{R}_+$ be $\mathcal{D}$-measurable, and let $h : D \mapsto E$ be measurable with respect to $\mathcal{D}$ and $\mathcal{E}$. Define

$$L(\omega, A) = \int_D N(\omega, dx)\ j(x)\ 1_A \circ h(x), \quad \omega \in \Omega, \quad A \in \mathcal{E}.$$

Show that L is an additive random measure on $(E, \mathcal{E})$.

4.33 *Dirichlet random measures.* In Lemma 4.3, suppose that μ is a finite measure on E and λ is the measure on $\mathbb{R}_+$ given by $\lambda(dz) = dz(e^{-z}/z)$. Then, L is a gamma random measure on E with $L(E) < \infty$ almost surely; see Example 4.12. Define

$$P(\omega, A) = \frac{L(\omega, A)}{L(\omega, E)}, \quad \omega \in \Omega, A \in \mathcal{E}.$$

Then, P is a random probability measure on $(E, \mathcal{E})$. We call it a Dirichlet random measure with shape measure μ.

The name comes from the Dirichlet distribution, which is a generalization of the beta distribution. For every finite measurable partition $\{A, \ldots, B\}$ of E, the vector $(P(A), \ldots, P(B))$ has the Dirichlet distribution with shape vector $(a, \ldots, b)$ where $a = \mu(A), \ldots, b = \mu(B)$. This distribution has the density function

$$\frac{\Gamma(a + \cdots + b)}{\Gamma(a) \cdots \Gamma(b)} x^{a-1} \cdots y^{b-1}, \quad (x, \ldots, y) \in \Delta,$$

where Δ is the simplex (of appropriate dimension) of positive vectors $(x, \ldots, y)$ with $x + \cdots + y = 1$.

In the particular case $a = \cdots = b = 1$, the distribution becomes the uniform distribution on Δ.

4.34 *Poisson-Dirichlet process.* Let L and P be as in the preceding exercise. Note that $c = \mu(E) < \infty$ and that $\lambda(z, \infty) < \infty$ for every $z > 0$. Thus the atoms of L can be labeled as $(Y_n, Z_n), n = 1, 2, \ldots$, so that $Z_1 > Z_2 > \ldots$. The sequence (Z_n) is called the Poisson-Dirichlet process in the statistical literature.

a) Show that (Z_n) forms a Poisson random measure on $\mathbb{R}_+$ with mean measure $c\lambda$.

b) Show that (Y_n) is independent of (Z_n) and is an independency of E-valued random variables with distribution $\frac{1}{c}\mu$.

c) Show that the Dirichlet random measure P has the form, with $S = \sum_1^\infty Z_n$,

$$P(A) = \frac{1}{S} \sum_{n=1}^{\infty} Z_n\ I(Y_n, A), \quad A \in \mathcal{E}.$$

4.35 *Sampling from Dirichlet.* Let P be a Dirichlet random measure on $(E, \mathcal{E})$ with shape measure μ. A collection $\{X_1, \ldots, X_n\}$ of E-valued random

variables is called a sample from P if, given P, the conditional law of $\{X_1, \ldots, X_n\}$ is that of an independency with the common distribution P, that is,

$$\mathbb{P}\{X_1 \in A_1, \ldots, X_n \in A_n | P\} = P(A_1) \ldots P(A_n), \quad A_1, \ldots, A_n \in \mathcal{E}.$$

Let $\{X_1, \ldots, X_n\}$ be such a sample. Show that, given $\{X_1, \ldots, X_n\}$, the conditional law of P is that of a Dirichlet random measure with shape measure

$$\mu_n(A) = \mu(A) + \sum_{i=1}^{n} I(X_i, A), \quad A \in \mathcal{E}.$$

Show that, assuming $\{X_1, \ldots, X_n, X_{n+1}\}$ is also a sample from P,

$$\mathbb{P}\{X_{n+1} \in A | X_1, \ldots, X_n\} = \frac{\mu_n(A)}{\mu_n(E)}, \quad A \in \mathcal{E}.$$

4.36 *Random fields.* By a positive random field on $(E, \mathcal{E})$ we mean a collection $F = \{F(x) : x \in E\}$ of positive random variables $F(x)$ such that the mapping $(\omega, x) \mapsto F(\omega, x)$ is measurable relative to $\mathcal{H} \otimes \mathcal{E}$ and $\mathcal{B}(\mathbb{R}_+)$. The probability law of F is specified by giving the finite-dimensional distributions, that is, the distribution of $(F(x_1), \ldots, F(x_n))$ with $n \geq 1$ and $x_1, \ldots, x_n$ in E. Equivalently, the probability law is specified by the Laplace transforms

$$\mathbb{E}\ e^{-\alpha F} = \mathbb{E}\ \exp_- \int_E \alpha(dx) F(x)$$

as α varies over all finite measures on $(E, \mathcal{E})$.

An expedient method of defining a positive random field is as follows: Let N be a Poisson random measure on some measurable space $(D, \mathcal{D})$, and let $k : D \times E \mapsto \mathbb{R}_+$ be $\mathcal{D} \otimes \mathcal{E}$-measurable. Define

$$F(\omega, y) = \int_D N(\omega, dx)\ k(x, y), \quad \omega \in \Omega, \quad y \in E.$$

Show that F is a positive random field on $(E, \mathcal{E})$. Show that, for every finite measure α on E, with ν the mean of N,

$$\mathbb{E}\ e^{-\alpha F} = \exp_- \int_D \nu(dx) \left[1 - \exp_- \int_E \alpha(dy)\ k(x, y)\right].$$

4.37 *Continuation.* Suppose that $E = \mathbb{R}^d$, $D = \mathbb{R}^d \times \mathbb{R}_+$, $\nu = \text{Leb} \times \lambda$ for some measure λ on $\mathbb{R}_+$, and take

$$k(x, r, y) = r\ 1_{\mathbb{B}}(y - x), \quad (x, r) \in D, \quad y \in E,$$

where $\mathbb{B}$ is the unit ball in $\mathbb{R}^d$ centered at the origin. Give a condition on λ to make F real-valued. For $d = 1$ and $\lambda(dz) = dz\ e^{-z}/z$, compute the marginal distribution of $F(y)$.

4.38 *Random vector fields.* By a *velocity field* on $\mathbb{R}^d$ is meant a mapping $u : \mathbb{R}^d \mapsto \mathbb{R}^d$, where one interprets $u(x)$ as the velocity of the particle at x. We now describe a random version, which is useful modeling the medium scale eddy motions in $\mathbb{R}^2$.

Let v be a smooth velocity field on the unit disk $D = \{x \in \mathbb{R}^2 : |x| \leq 1\}$. Let N be a Poisson random measure on $E = \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+$ with mean $\nu = \text{Leb} \times \alpha \times \beta$ where α is a probability measure on $\mathbb{R}$, and β on $\mathbb{R}_+$. Define

$$u(\omega, x) = \int_E N(\omega; dz, da, db)\ a\ v(\frac{x-z}{b}), \quad \omega \in \Omega, \quad x \in \mathbb{R}^2.$$

This defines a two-dimensional random velocity field u on $\mathbb{R}^2$.

Think of v as the velocity field corresponding to an eddy motion over the unit disk D. Then, $x \mapsto v(\frac{x-z}{b})$ corresponds to a similar eddy motion over a disk of radius b centered at z. Thus, $u(\omega, \cdot)$ is the superposition of velocity fields $x \mapsto av(\frac{x-z}{b})$, each one corresponding to an atom (z, a, b) of the counting measure $N(\omega, \cdot)$.

Show that μ is homogeneous, that is, its probability law is invariant under shifts of the origin. If v is isotropic, show that the probability law of u is invariant under rotations as well.

5 Poisson Processes

This section is devoted to a closer examination of simple Poisson processes and, by extension, of Poisson random measures on $\mathbb{R}_+$. The presence of time and the order properties of the real-line allow for a deeper understanding of such processes. This, in turn, illustrates the unique position occupied by things Poisson in the theories of martingales, point processes, and Markov processes.

Counting processes

We introduce the setup to be kept throughout this section. In keeping with traditional notation, we let $N = (N_t)_{t \in \mathbb{R}_+}$ be a counting process: for almost every ω, the path $t \mapsto N_t(\omega)$ is an increasing right-continuous step function with $N_0(\omega) = 0$ and whose every jump is of size one. Such a process is defined by its jump times: there is an increasing sequence of random variables T_k taking values in $\bar{\mathbb{R}}_+$ such that

5.1 $$N_t(\omega) = \sum_{k=1}^{\infty} 1_{[0,t]} \circ T_k(\omega), \quad t \in \mathbb{R}_+,\ \omega \in \Omega.$$

It is usual to call (T_k) the point process associated with N. Since $N_0 = 0$ and each jump of N is of size one,

5.2 $$0 < T_1(\omega) < T_2(\omega) < \cdots < T_k(\omega) \quad \text{if } T_k(\omega) < \infty,$$

this being true for almost every ω. Of course, if $T_k(\omega) = \infty$ then $T_n(\omega) = \infty$ for all $n > k$. The sequence (T_k) forms a random counting measure M on $\mathbb{R}_+$; for positive Borel functions f on $\mathbb{R}_+$, extending f onto $\bar{\mathbb{R}}_+$ by setting $f(\infty) = 0$,

5.3 $$Mf = \sum_{k=1}^{\infty} f \circ T_k.$$

Obviously, $N_t(\omega) = M(\omega, [0,t])$. Finally, we let $\mathcal{F}$ be the filtration generated by N, that is,

5.4 $$\mathcal{F}_t = \sigma\{N_s : s \le t\} = \sigma\{M(A) : A \in \mathcal{B}_{[0,t]}\}.$$

Poisson processes

We recall Definition V.2.20 in the present context: The counting process N is called a *Poisson process* with rate c if, for every s and t in $\mathbb{R}_+$, the increment $N_{s+t} - N_s$ is independent of $\mathcal{F}_s$ and has the Poisson distribution with mean ct. Equivalently, N is Poisson with rate c if and only if it has stationary and independent increments (see Definition 4.5a,b) and each N_t has the Poisson distribution with mean ct.

The following theorem characterizes Poisson processes in terms of random measures, Markov processes, martingales, and point processes. Much of the proof is easy and is of a review nature. The exceptions are the proofs that (c)⇒(d)⇒(a).

5.5 THEOREM. *For fixed c in $(0, \infty)$, the following are equivalent:*

a) M is a Poisson random measure with mean $\mu = c$ Leb.
b) N is a Poisson counting process with rate c.
c) N is a counting process and $\tilde{N} = (N_t - ct)_{t \in \mathbb{R}_+}$ is an $\mathcal{F}$-martingale.
d) (T_k) is an increasing sequence of $\mathcal{F}$-stopping times, and the differences $T_1, T_2 - T_1, T_3 - T_2, \ldots$ are independent and exponentially distributed with parameter c.

REMARK. The limitation on c is natural: Since N is a counting process, N_t is finite almost surely and can have the Poisson distribution with mean ct only if c is finite. The other possibility, $c = 0$, is without interest, because it implies that $N = 0$ almost surely.

Proof. We shall show that (a)⇒(b) $\iff$ (c)⇒(d)⇒(a).

i) Assume (a). Since the Lebesgue measure is diffuse, M is a random counting measure by Theorem 2.14, and $\mu[0,t] = ct < \infty$. Thus, N is a counting process. The independence of $N_{s+t} - N_s$ from $\mathcal{F}_s$ and the associated distribution being Poisson with mean ct follow from the definition of Poisson random measures. So, (a)⇒(b).

ii) The equivalence (b) $\iff$ (c) was shown in Chapter V on martingales; see Theorem V.2.23 and Proposition V.6.13.

iii) Assume (c) and, therefore, (b). It follows from 5.1 that, for each integer $k \geq 1$, we have $\{T_k \leq t\} = \{N_t \geq k\} \in \mathcal{F}_t$ for every t. Thus, each T_k is a stopping time of $\mathcal{F}$. Moreover, since N is a counting process and $\lim_{t\to\infty} N_t = M(\mathbb{R}_+) = +\infty$ almost surely, for almost every ω, we have 5.2, and $T_k(\omega) < \infty$ for all k, and $\lim_k T_k(\omega) = \infty$.

On the other hand, Corollary V.6.7 applies to the martingale $\tilde{N}$: we have, with $\mathbb{E}_S = \mathbb{E}(\cdot|\mathcal{F}_S)$ as usual,

5.6 $$\mathbb{E}_S \int_{(S,T]} F_t \; dN_t = \mathbb{E}_S \int_{(S,T]} F_t \; c \; dt$$

for bounded predictable processes F and stopping times S and T with $S \leq T$. Put $T_0 = 0$ for convenience, take $S = T_k$ and $T = T_{k+1}$ for fixed k in $\mathbb{N}$, and let $F_t = re^{-rt}$ with $r > 0$ fixed. On the left side of 5.6, the integral becomes equal to re^{-rT} since N remains constant over the interval (S, T) and jumps by the amount one at T. So, 5.6 becomes

$$r \; \mathbb{E}_S e^{-rT} = c \; \mathbb{E}_S(e^{-rS} - e^{-rT}).$$

Multiplying both sides by e^{rS}, which can be passed inside the conditional expectations since S is in $\mathcal{F}_S$, and re-arranging, we obtain

$$\mathbb{E}_S e^{-r(T-S)} = \frac{c}{c+r}, \quad r > 0.$$

This means that $T - S = T_{k+1} - T_k$ is independent of $\mathcal{F}_S = \mathcal{F}_{T_k}$ and has the exponential distribution with parameter c. Thus (c)⇒(d).

iv) Assume (d). Let $\hat{M}$ be a Poisson random measure on $\mathbb{R}_+$ with mean $\hat{\mu} = \mu = c$ Leb, constructed over some auxiliary probability space $(\hat{\Omega}, \hat{\mathcal{H}}, \hat{\mathbb{P}})$; see Theorem 2.13 for its existence. Using the already proved implications (a)⇒(b)⇒(c)⇒(d) on the Poisson random measure $\hat{M}$, we conclude that, for positive Borel f,

5.7 $$\hat{M}f = \sum_1^\infty f \circ \hat{T}_k,$$

where $\hat{T}_1, \hat{T}_2 - \hat{T}_1, \hat{T}_3 - \hat{T}_2, \ldots$ are independent and have the exponential distribution with parameter c. Observe: 5.3 and 5.7 hold, and (T_k) and $(\hat{T}_k)$ have the same probability law. It follows that the random variables Mf and $\hat{M}f$ have the same distribution. Hence, writing $\hat{\mathbb{E}}$ for expectation on $(\hat{\Omega}, \hat{\mathcal{H}}, \hat{\mathbb{P}})$,

$$\mathbb{E} \; e^{-Mf} = \hat{\mathbb{E}} \; e^{-\hat{M}f} = e^{-\mu(1-e^{-f})},$$

where the last equality follows from Theorem 2.9 on Laplace functionals applied to the Poisson random measure $\hat{M}$ with mean $\hat{\mu} = \mu$. This shows, via Theorem 2.9, that M is Poisson with mean μ. Thus, (d)⇒(a). □

5.8 REMARK. The unorthodox proof that (d)⇒(a) is an application of the following principle. Let X and Y be random variables taking values on some spaces, and suppose that they define each other, that is, $Y = h \circ X$ for some isomorphism h and $X = g \circ Y$ with g being the functional inverse of h. Let λ be the distribution of X and let $\mu = \lambda \circ h^{-1}$. The principle, rather trivial, is that Y has the distribution μ if and only if X has the distribution λ. The proof of (d)⇒(a) uses this principle with $X = M$, $Y = (T_k)$, and λ the probability law of M. Unfortunately, a direct application of the principle requires showing that h is a bimeasurable bijection, which is technically difficult: in this application, h is a mapping from the space of all counting measures on $\mathbb{R}_+$ onto the space of all increasing sequences on $\mathbb{R}_+$. The auxiliary process $\hat{M}$ was introduced to circumvent the technical difficulties.

Characterization as a Lévy process

The next theorem's characterization is often used as a definition: A Poisson process is a counting process with stationary and independent increments. We refer to Definition 4.5 for increasing Lévy processes and the allied terminology.

5.9 THEOREM. *The counting process N is a Lévy process if and only if it is a Poisson process.*

Proof. Sufficiency part is trivial. As to the necessity part, assuming that N is a counting Lévy process, the only thing to show is that N_t has the Poisson distribution with mean ct for some constant c in $\mathbb{R}_+$ and for all times t.

a) We start by showing that, for some fixed c in $\mathbb{R}_+$,

5.10 $$q(t) = \mathbb{P}\{N_t = 0\} = e^{-ct}, \quad t \in \mathbb{R}_+.$$

For s and t in $\mathbb{R}_+$, because N is Lévy, the increments N_s and $N_{s+t} - N_s$ are independent and the latter has the same distribution as N_t. Thus,

$$\mathbb{P}\{N_{s+t} = 0\} = \mathbb{P}\{N_s = 0, N_{s+t} - N_s = 0\} = \mathbb{P}\{N_s = 0\}\mathbb{P}\{N_t = 0\},$$

that is,

5.11 $$q(s+t) = q(s)q(t), \quad s, t \in \mathbb{R}_+.$$

Moreover, $q(0) = 1$, and $q(t) = \mathbb{E}\, 1_{\{0\}} \circ N_t$ is right-continuous by the almost sure right-continuity of $t \mapsto 1_{\{0\}} \circ N_t$ and the bounded convergence theorem for expectations. The only solution of 5.11 with these properties is as claimed in 5.10 with c in $\mathbb{R}_+$. If $c = 0$, then $N_t = 0$ almost surely for all t, which makes N a trivial Poisson process, and the proof is complete. For the remainder of the proof we assume that $0 < c < \infty$.

b) It is convenient to recall M, the random counting measure associated with $N = (N_t)$. Fix $t > 0$. Let $A_1, \dots, A_n$ be equi-length intervals of the form $(\cdot, \cdot]$ that constitute a partition of $A = (0, t]$. Let X_i be the indicator of the event $\{M(A_i) \geq 1\}$ and put $S_n = X_1 + \cdots + X_n$. Since N is Lévy, the increments $M(A_1), \dots, M(A_n)$ are independent and have the same distribution as $M(A_1) = N_{t/n}$. Thus, $X_1, \dots, X_n$ are independent Bernoulli variables with the same success probability $p = 1 - q(t/n) = 1 - e^{-ct/n}$. Hence, for each k in $\mathbb{N}$,

$$\begin{aligned}\mathbb{P}\{S_n = k\} &= \frac{n!}{k!(n-k)!}(1 - e^{-ct/n})^k (e^{-ct/n})^{n-k} \\ &= \frac{e^{-ct}}{k!} n(n-1)\cdots(n-k+1)(e^{ct/n} - 1)^k.\end{aligned}$$

c) For almost every ω, the counting measure M_ω has only finitely many atoms in A; let $\delta(\omega)$ be the minimum distance between the atoms; if $n > t/\delta(\omega)$ then we have $S_n(\omega) = M(\omega, A) = N_t(\omega)$. In other words, $S_n \to N_t$ almost surely as $n \to \infty$. Hence,

$$\mathbb{P}\{N_t = k\} = \lim_{n\to\infty} \mathbb{P}\{S_n = k\} = \frac{e^{-ct}(ct)^k}{k!}, \quad k \in \mathbb{N}.$$

This completes the proof that N is a Poisson process. □

The preceding proof is the most elementary of all possible ones. In the next chapter we shall give another proof. Assuming that N is a counting Lévy, we shall show that $\mathbb{E}\, N_t = ct$ necessarily for some finite constant c; this uses the strong Markov property (to be shown). Then, $(N_t - ct)$ is a martingale, and N is Poisson by Theorem 5.5 or, more correctly, by Proposition V.6.13.

A minimalist characterization

A careful examination of the preceding proof would show that the Lévy property of N is used only to conclude that the Bernoulli variables X_i are independent and have the same success probability. This observation leads to the following characterization theorem; compare the minimal nature of the condition here with the extensive conditions of Definition 2.3. See 5.16 below for a generalization.

5.12 THEOREM. *Let M be a random counting measure on $\mathbb{R}_+$. Let $\mu = c$ Leb on $\mathbb{R}_+$ for some constant c in $\mathbb{R}_+$. Then M is Poisson with mean μ if and only if*

5.13 $$\mathbb{P}\{M(A) = 0\} = e^{-\mu(A)}$$

for every bounded set A that is a finite union of disjoint intervals.

Proof. The necessity of the condition follows trivially from Definition 2.3. To prove the sufficiency, assume that 5.13 holds for every A as described, and let $\mathcal{A}$ be the collection of all such A.

Fix a finite number of disjoint intervals $A, \ldots, B$; let $a, \ldots, b$ be the corresponding lengths; and let $i, \ldots, j$ be positive integers. We shall show that

5.14 $$\mathbb{P}\{M(A) = i, \ldots, M(B) = j\} = \frac{e^{-ca}(ca)^i}{i!} \cdots \frac{e^{-cb}(cb)^j}{j!}.$$

This will prove that N is a Poisson process with rate c and, hence, that M is Poisson with mean μ as claimed.

a) Let $A_1, \ldots, A_n$ form a partition of A into n equi-length sub-intervals, $\ldots$, and $B_1, \ldots, B_n$ form a partition of B similarly. Let

$$\mathcal{D} = \{A_1, \ldots, A_n; \ldots; B_1, \ldots, B_n\}.$$

For each D in $\mathcal{D}$, let X_D be the indicator of $\{N(D) \geq 1\}$, and define

5.15 $$S_n(A) = \sum_1^n X_{A_m}, \ldots, S_n(B) = \sum_1^n X_{B_m}.$$

Arguments of the proof of Theorem 5.9, part (c), show that

$$M(A) = \lim_n S_n(A), \ldots, M(B) = \lim_n S_n(B)$$

almost surely. Thus, to prove 5.14, it is enough to show that $S_n(A), \ldots, S_n(B)$ are independent and have the binomial distributions for n trials with respective success probabilities $1 - e^{-ca/n}, \ldots, 1 - e^{-cb/n}$.

b) Consider the collection $\{X_D : D \in \mathcal{D}\}$ of Bernoulli variables. Let $\mathcal{C}$ be a subset of $\mathcal{D}$, and let C be the union of the elements of $\mathcal{C}$. Observe that $1 - X_D$ is the indicator of $\{M(D) = 0\}$, and the product $\prod(1 - X_D)$ over D in $\mathcal{C}$ is the indicator of $\{M(C) = 0\}$. Since $C \in \mathcal{A}$ and $\mathcal{D} \subset \mathcal{A}$, the condition 5.13 holds for the sets D and C; thus,

$$\mathbb{E} \prod_{D \in \mathcal{C}} (1 - X_D) = e^{-\mu(C)} = \prod_{D \in \mathcal{C}} e^{-\mu(D)} = \prod_{D \in \mathcal{C}} \mathbb{E}\,(1 - X_D).$$

This implies, via II.5.33, that the collection $\{1 - X_D : D \in \mathcal{D}\}$ is an independency. Hence, $\{X_D : D \in \mathcal{D}\}$ is an independency, and the success probabilities are

$$\mathbb{P}\{X_D = 1\} = 1 - e^{-\mu(D)},$$

where $\mu(D)$ is equal to ca/n for the first n elements of $\mathcal{D}$, and $\ldots$, and to cb/n for the last n elements. Thus, $S_n(A), \ldots, S_n(B)$ are independent and have the binomial distribution claimed in part (a). This completes the proof. □

The preceding can be generalized to abstract spaces. The basic idea is the following: Let $(E, \mathcal{E})$ be a standard measurable space and let μ be a diffuse Σ-finite measure on it. Put $b = \mu(E) \le +\infty$. As was explained in Exercise I.5.16, it is possible to find measurable bijections $g : E \mapsto [0, b)$ and $h : [0, b) \mapsto E$ such that $\hat{\mu} = \mu \circ g^{-1}$ is the Lebesgue measure on $[0, b)$, and $\mu = \hat{\mu} \circ h^{-1}$ in return.

5.16 THEOREM. *Let μ be a diffuse Σ-finite measure on a standard measurable space $(E, \mathcal{E})$. Let M be a random counting measure on E. Suppose that*

5.17 $$\mathbb{P}\{M(A) = 0\} = e^{-\mu(A)}, \quad A \in \mathcal{E}.$$

Then, M is a Poisson random measure with mean μ.

Proof. Let g and h be as described preceding the theorem. Since g is a bijection, $\hat{M} = M \circ g^{-1}$ is still a random counting measure, and, with $\hat{\mu} = \mu \circ g^{-1} =$ Lebesgue on $[0, b)$, we have

$$\mathbb{P}\{\hat{M}(B) = 0\} = \mathbb{P}\{M(g^{-1}B) = 0\} = e^{-\mu(g^{-1}B)} = e^{-\hat{\mu}(B)}$$

for every Borel subset of $[0, b)$. Since $\hat{\mu}$ is the Lebesgue measure on $[0, b)$, Theorem 5.12 applies to conclude that $\hat{M}$ is Poisson on $[0, b)$ with mean $\hat{\mu}$. It follows that $M = \hat{M} \circ h^{-1}$ is Poisson on E with mean $\hat{\mu} \circ h^{-1} = \mu$. □

Strong Markov property

As was remarked in V.6.16, if N is a Poisson process with rate c, then the independence of increments can be extended to increments over random intervals $(S, S + t]$: For every finite stopping time S of the filtration $\mathcal{F}$, the increment $N_{S+t} - N_S$ is independent of $\mathcal{F}_S$ and has the Poisson distribution with mean ct. We may remove the finiteness condition on S and express the same statement as

5.18 $$\mathbb{E}_S f(N_{S+t} - N_S)\, 1_{\{S<\infty\}} = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} f(k) 1_{\{S<\infty\}}.$$

In fact, we could have used this result to shorten the proof of (c)⇒(d) in Theorem 5.5: take $f = 1_{\{0\}}$ and $S = T_k$ to conclude that the event $\{N_{T_k+t} - N_{T_k} = 0\} = \{T_{k+1} - T_k > t\}$ is independent of $\mathcal{F}_{T_k}$ and has probability e^{-ct}.

The property 5.18 is called the strong Markov property for the Poisson process N. It expresses the independence of future from the past when the present is a stopping time.

5.19 EXAMPLE. Let S be the first time an interval of length a passes without a jump, that is,

$$S = \inf\{t \ge a : N_t = N_{t-a}\}.$$

Clearly, $S = T_k + a$ if and only if the first k interjump intervals are at most a in length and the $(k+1)^{\text{th}}$ interval is greater than a in length. Thus, $S < \infty$ almost surely. Let T be the time of first jump after S. Note that the interval that includes S has length $a+(T-S)$, and for a large, the raw intuition expects $T-S$ to be small. Instead, noting that $\{T - S > t\} = \{N_{S+t} - N_S = 0\}$, we see that $T-S$ is independent of $\mathcal{F}_S$ and has the same exponential distribution as if S is a jump time.

Total unpredictability of jumps

Let $T = T_1$ be the time of first jump for the Poisson process N with rate c. Is it possible to predict T? Is there a sequence of stopping times S_n increasing to T and having $S_n < T$ almost surely? The following shows that the answer is no.

5.20 PROPOSITION. *Let S be a stopping time of $\mathcal{F}$. Suppose that $0 \le S < T$ almost surely. Then, $S = 0$ almost surely.*

Proof. Since $S < T$, the event $\{T - S > t\}$ is the same as the event $\{N_{S+t} - N_S = 0\}$, and the latter is independent of $\mathcal{F}_S$ and has probability e^{-ct} by the strong Markov property 5.18. Thus, in particular,

$$\mathbb{E}_S(T - S) = 1/c.$$

Taking expectations, and recalling that $\mathbb{E}\,T = 1/c$, we conclude that $\mathbb{E}\,S = 0$. Thus, $S = 0$ almost surely. □

A similar proof will show that if $T_k \le S < T_{k+1}$ almost surely then $S = T_k$ almost surely, this is for each k. Thus, it seems impossible to predict the jump time T_k for fixed k. We list the following stronger result without proof. See V.7.30 and V.7.31 for the terms. Recall the definition 5.4 of $\mathcal{F}$.

5.21 THEOREM. *Let S be a stopping time of $\mathcal{F}$. Then, S is predictable if and only if $N_{S-} = N_S$ almost surely on $\{S > 0\}$; and S is totally unpredictable if and only if $N_{S-} \ne N_S$ almost surely.*

Exercises

5.22 *Crossing times.* Let S be as in Example 5.19. Find its expected value.

5.23 *Logarithmic Poisson random measures.* Let $U_1, U_2, \ldots$ be independent and uniformly distributed over $E = (0,1)$. Let $X_n = U_1 U_2 \cdots U_n$, $n \ge 1$, and define M to be the random counting measure on E whose atoms are those X_n. Show that M is a Poisson random measure on E.

5.24 *Continuation.* Let M be a Poisson random measure on $E = (0, \infty)$ with mean μ given by

$$\mu(dx) = dx\,\frac{1}{x}, \quad x \in E.$$

Note that $M(A) < \infty$ almost surely for every closed interval $A = [a, b]$ in E, but that $M(A) = +\infty$ almost surely for $A = (0, a)$ and for $A = (b, \infty)$. Label the atoms X_i such that $\cdots < X_{-1} < X_0 < 1 \le X_1 < X_2 < \cdots$. Describe the probability law of (X_i).

5.25 *Atom counts.* Let L be a purely atomic additive random measure on $\mathbb{R}_+$. Suppose that

$$\mathbb{P}\{L(A) = 0\} = e^{-\text{Leb } A}, \quad A \in \mathcal{B}_{\mathbb{R}_+}.$$

Describe the atomic structure of L.

5.26 *Importance of additivity.* Let (T_k) be the increasing sequence of jump times for a Poisson process (N_t) with unit rate. Let $w_1, w_2, \ldots$ be arbitrarily chosen from $(0, \infty)$. Define

$$L(A) = \sum_{i=1}^{\infty} w_i \; I(T_i, A), \quad A \in \mathcal{B}_{\mathbb{R}_+}.$$

Show that $\mathbb{P}\{L(A) = 0\} = e^{-\text{Leb } A}$. Is the random measure L additive?

5.27 *Another warning.* Let $(N_t)_{t \in \mathbb{R}_+}$ be a counting process. Suppose that N_t has the Poisson distribution with mean t for every t in $\mathbb{R}_+$. Give an example where N is *not* a Poisson process.

5.28 *Importance of diffusivity.* In Theorem 5.16, the diffusivity of μ is essential. On $\mathbb{R}$, define $\mu = c\delta_0$ and $M = K\delta_0$, where $c > 0$ is a constant, and K is equal to 0 or 2 with respective probabilities e^{-c} and $1 - e^{-c}$. Show that M is not Poisson.

6 Poisson Integrals and Self-exciting Processes

Our aim is the to introduce some martingale-theoretic tools in dealing with random measures. Some such tools were introduced in Chapter V on martingales and were used in the martingale characterization for Poisson processes. Here, we introduce similar ideas for applications to stochastic processes with jumps, because the jumps are often regulated by Poisson random measures. As an example, we introduce counting processes with random intensities and a particular self-exciting two point process. Further applications will appear in the chapters on Lévy and Markov processes.

Throughout, $(E, \mathcal{E})$ will be a measurable space, and $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ an augmented filtration on the probability space $(\Omega, \mathcal{H}, \mathbb{P})$.

Poisson on time and space

This is about Poisson random measures on spaces of the form $\mathbb{R}_+ \times E$, where $\mathbb{R}_+$ is interpreted as time and E as some physical space, and the filtration $\mathcal{F}$ represents the flow of information over time. As usual $\mathbb{R}_+ \times E$ is

furnished with the product σ-algebra $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{E}$, and we shall omit the mention of the σ-algebra in phrases like "random measure on $\mathbb{R}_+ \times E$."

6.1 Definition. *Let M be a Poisson random measure on $\mathbb{R}_+ \times E$. It is said to be so relative to the filtration $\mathcal{F}$ if, for every t in $\mathbb{R}_+$,*

a) $M(A)$ is in $\mathcal{F}_t$ for every A in $\mathcal{B}_{[0,t]} \otimes \mathcal{E}$, and
b) the trace of M over $(t, \infty) \times E$ is independent of $\mathcal{F}_t$.

The condition (a) is called adaptedness, and (b) the independence of the future of M from the past. It is obvious that, if M is Poisson relative to $\mathcal{F}$, then it is Poisson relative to the filtration $\mathcal{G}$ generated by itself (that is, $\mathcal{G}_t$ is the σ-algebra generated by $M(A)$, $A \in \mathcal{B}_{[0,t]} \otimes \mathcal{E}$).

Given a Poisson random measure M on $\mathbb{R}_+ \times E$, it is convenient to think of its atoms as solid objects. For a fixed outcome ω, if (t, z) is an atom of the measure M_ω, then we think of t as the arrival time of that atom and of z as the landing point in space; see also Proposition 2.18 where $E = \mathbb{R}_+$ and is interpreted as "size" space.

Poisson integrals of predictable processes

This is to extend Theorem V.6.5 to the current case of processes on $\mathbb{R}_+ \times E$. Recall Definition 6.1 and Proposition 6.2 from Chapter V: The $\mathcal{F}$-predictable σ-algebra is the σ-algebra $\mathcal{F}^p$ on $\Omega \times \mathbb{R}_+$ generated by the collection of sets having the form $H \times A$, where A is an interval $(a, b]$ and H is an event in $\mathcal{F}_a$, or A is the singleton $\{0\}$ and H is in $\mathcal{F}_0$. A process $F = (F_t)_{t \in \mathbb{R}_+}$ is said to be $\mathcal{F}$-predictable, or is said to be in $\mathcal{F}^p$, if the mapping $(\omega, t) \mapsto F_t(\omega)$ from $\Omega \times \mathbb{R}_+$ into $\mathbb{R}$ is $\mathcal{F}^p$-measurable. In particular, if F is adapted to $\mathcal{F}$ and is left-continuous, then it is $\mathcal{F}$-predictable. The next theorem is important.

6.2 Theorem. *Let M be a Poisson random measure on $\mathbb{R}_+ \times E$ with mean measure μ satisfying $\mu(\{0\} \times E) = 0$. Let $G = \{G(t, z) : t \in \mathbb{R}_+, z \in E\}$ be a positive process in $\mathcal{F}^p \otimes \mathcal{E}$. Suppose that M is Poisson relative to $\mathcal{F}$. Then,*

6.3 $$\mathbb{E} \int_{\mathbb{R}_+ \times E} M(dt, dz)\ G(t, z) = \mathbb{E} \int_{\mathbb{R}_+ \times E} \mu(dt, dz)\ G(t, z).$$

Proof. The hypothesis on G is that the mapping $(\omega, t, z) \mapsto G(\omega, t, z)$ from $\Omega \times \mathbb{R}_+ \times E$ into $\bar{\mathbb{R}}_+$ is $(\mathcal{F}^p \otimes \mathcal{E})$-measurable. Consider the collection of all such G for which 6.3 holds. That collection includes constants, is a linear space, and is closed under increasing limits, the last being by the monotone convergence theorem applied to 6.3 with $G_n \nearrow G$ on both sides. Thus, by the monotone class theorem, the proof is reduced to showing 6.3 for G that are indicators of sets $H \times A \times B$, where $B \in \mathcal{E}$ and $A = (a, b]$ and $H \in \mathcal{F}_a$,

and also when $B \in \mathcal{E}$ and $A = \{0\}$ and $H \in \mathcal{F}_0$. In the former case, the left side of 6.3 is equal to

$$\begin{aligned}\mathbb{E}\ 1_H\ M(A \times B) &= \mathbb{E}\ \mathbb{E}_a\ 1_H\ M(A \times B) \\ &= \mathbb{E}\ 1_H\ \mathbb{E}_a\ M(A \times B) \\ &= \mathbb{E}\ 1_H\ \mathbb{E}\ M(A \times B) = \mathbb{E}\ 1_H\ \mu(A \times B),\end{aligned}$$

and the last member is exactly the right side of 6.3; here we used the notation $\mathbb{E}_a$ for $\mathbb{E}(\cdot|\mathcal{F}_a)$ as in earlier chapters and noted that $\mathbb{E} = \mathbb{E}\ \mathbb{E}_a$, and $H \in \mathcal{F}_a$, and $A \times B \subset (a, \infty] \times E$, and the trace of M over $(a, \infty) \times E$ is independent of $\mathcal{F}_a$. The case where $A = \{0\}$ is trivial: then, $\mu(A \times E) = 0$ by hypothesis, which implies that $M(A \times E) = 0$ almost surely, which together imply that the integrals vanish on both sides of 6.3. □

6.4 Remarks. a) In most applications the mean μ will have the form $\mu = \text{Leb} \times \lambda$ for some measure λ on E. Then, $\mu(\{0\} \times E) = 0$ automatically.

b) Let S and T be stopping times of $\mathcal{F}$ with $S \le T$, and let V be a positive random variable in $\mathcal{F}_S$. Then, putting $F_t = V\ 1_{(S,T]}(t)$ for t in $\mathbb{R}_+$, we obtain a positive predictable (in fact, left-continuous) process F. Letting M and G be as in the preceding theorem, we note that $\tilde{G}(t, z) = F_t G(t, z)$ defines a process $\tilde{G}$ that is again positive and in $\mathcal{F}^p \otimes \mathcal{E}$. Thus, replacing G by $\tilde{G}$ in 6.3, we see that

$$\mathbb{E}_S \int_{(\!(S,T]\!] \times E} M(dt, dz)\ G(t, z) = \mathbb{E}_S \int_{(\!(S,T]\!] \times E} \mu(dt, dz)\ G(t, z),$$

where $(\!(S, T]\!]$ should be interpreted as $(S, T] \cap \mathbb{R}_+$ in order to accommodate possibly infinite values for S and T.

c) Let M and G be as in the theorem. Suppose that

$$\mathbb{E} \int_{[0,t] \times E} \mu(ds, dz)\ G(s, z) < \infty, \quad t \in \mathbb{R}_+.$$

Then,

$$X_t = \int_{[0,t] \times E} M(ds, dz)\ G(s, z) - \int_{[0,t] \times E} \mu(ds, dz)\ G(s, z)$$

is integrable for each t, and is in $\mathcal{F}_t$ for each t. The preceding remark applied with deterministic s and t show that $X = (X_t)_{t \in \mathbb{R}_+}$ is a martingale with $X_0 = 0$.

6.5 Remark. In fact, Theorem 6.2 has a partial converse that can be used to characterize Poisson random measures: Let M be a random counting measure on $\mathbb{R}_+ \times E$ whose mean μ is equal to $\text{Leb} \times \lambda$, where λ is a σ-finite measure on E. Suppose that M is adapted to $\mathcal{F}$, that is, 6.1a holds, and that 6.3 holds for every positive G in $\mathcal{F}^p \otimes \mathcal{E}$. Finally, suppose that no two atoms

arrive simultaneously, that is, for almost every ω, we have $M(\omega, \{t\} \times E) \le 1$ for all t. Then, it can be shown that M is a Poisson random measure relative to $\mathcal{F}$. Here is an outline of the proof.

For B in $\mathcal{E}$, put $N_t = M([0,t] \times B)$, and assume that $\lambda(B) < \infty$. Then, $\mathbb{E}\, N_t = \lambda(B)t$, and by the assumption that no two atoms of M arrive simultaneously, (N_t) is a counting process. In the formula 6.3 being assumed, taking $G(t,z) = F_t\, 1_B(z)$ for some positive predictable process F, we see that

$$\mathbb{E} \int_{\mathbb{R}_+} F_t\, dN_t = \lambda(B)\, \mathbb{E} \int_{\mathbb{R}_+} F_t\, dt.$$

Now, it follows from Proposition V.6.13 (see also V.6.3 - V.6.12) that (N_t) is a Poisson process with rate $\lambda(B)$. Further, it can be shown that the Poisson processes corresponding to disjoint sets $B_1, \ldots, B_n$ are independent, because no two such processes can jump at the same time; this will be proved in the next chapter. It follows that M is a Poisson random measure on $\mathbb{R}_+ \times E$.

Self-exciting processes

For a Poisson arrival process with constant rate c, the expected number of arrivals during an interval of length t is equal to ct, and, hence, c is the arrival intensity in this sense. In Exercise 2.36, the rate c was replaced by a random rate R_t varying with time, but the randomness of R stemmed from a source exogeneous to the arrival process N. Here, we extend the concept to cases where the rate R_t at time t is allowed to depend on the history of arrivals during $(0,t)$. Recall that $\mathcal{F} = (\mathcal{F}_t)$ is an arbitrary filtration over time. We use the shorthand notation $\mathbb{E}_t$ for the conditional expectation $\mathbb{E}_{\mathcal{F}_t} = \mathbb{E}(\cdot|\mathcal{F}_t)$.

6.6 Definition. *Let $N = (N_t)_{t\in\mathbb{R}_+}$ be a counting process adapted to $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Let $R = (R_t)_{t\in\mathbb{R}_+}$ be a positive $\mathcal{F}$-predictable process. Then, R is called the* intensity process *for N relative to $\mathcal{F}$ if*

6.7 $$\mathbb{E} \int_{\mathbb{R}_+} F_t\, dN_t = \mathbb{E} \int_{\mathbb{R}_+} F_t\, R_t\, dt$$

for every positive $\mathcal{F}$-predictable process F.

The processes F satisfying 6.7 form a positive monotone class. Noting that $N_0 = 0$ almost surely, we conclude that 6.7 holds for every positive predictable F if and only if it holds for those F that are the indicators of sets of the form $H \times (t,u]$ with $0 \le t < u$ and H in $\mathcal{F}_t$. Hence, the condition 6.7 is equivalent to the following:

6.8 $$\mathbb{E}_t\, (N_u - N_t) = \mathbb{E}_t \int_{(t,u]} ds\, R_s, \quad 0 \le t < u.$$

Heuristically, putting $u = t + \triangle t$ for small $\triangle t$, we may say that $R_t \triangle t$ is the conditional expectation of $N_{t+\triangle t} - N_t$ given $\mathcal{F}_t$; and, hence, the term

intensity at t for the random variable R_t. Also, especially when R_t depends on the past $\{N_s : s < t\}$, the process N is said to be *self-exciting*.

Another interpretation for 6.7 is to re-read 6.8 when $\mathbb{E}\, N_t < \infty$ for every t: Then,

6.9 $$\tilde{N}_t = N_t - \int_{(0,t]} ds\ R_s, \quad t \in \mathbb{R}_+,$$

is a martingale, and for every bounded predictable process F,

6.10 $$L_t = \int_{(0,t]} F_s\ d\tilde{N}_s, \quad t \in \mathbb{R}_+,$$

is again a martingale, both with respect to $\mathcal{F}$ obviously.

Given that the counting process N admits R as its intensity process, the condition 6.8 and Theorem V.6.18 show that N is a Poisson process relative to $\mathcal{F}$ if and only if R is deterministic, that is, $R_t = r(t)$ for some positive Borel function r on $\mathbb{R}_+$ with finite integral over bounded intervals. For the process N constructed in Exercise 2.36, where N is conditionally Poisson, R_t is random but is not affected by N, that is, R is exogeneous to N (of course, N depends on R). The following suggests a construction for N in all cases. Recall that $I(z, B)$ is one if $z \in B$ and is zero otherwise.

6.11 Theorem. *Let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $\mu = Leb \times Leb$, and suppose that it is Poisson relative to $\mathcal{F}$ in the sense of 6.1. Suppose that N is a counting process that satisfies*

6.12 $$N_t(\omega) = \int_{[0,t]\times\mathbb{R}_+} M(\omega; ds, dz)\ I(z, (0, R_s(\omega)]), \quad \omega \in \Omega, \quad t \in \mathbb{R}_+,$$

for some positive $\mathcal{F}$-predictable process R. Then, N is adapted to $\mathcal{F}$ and has R as its intensity relative to $\mathcal{F}$. Moreover, for every increasing function f on $\mathbb{N} = \{0, 1, \ldots\}$,

6.13 $$\mathbb{E}\ f\circ N_t = f(0) + \mathbb{E} \int_{[0,t]} ds\ \ [f\circ(N_s + 1) - f\circ N_s]\ R_s, \quad t \in \mathbb{R}_+.$$

6.14 Remark. In practice, 6.13 is more useful as a differential equation:

6.15 $$\frac{d}{dt}\ \mathbb{E}\ f\circ N_t = \mathbb{E}\ R_t\ \ [f\circ(N_t + 1) - f\circ N_t], \quad t > 0.$$

In particular, taking $f(n) = 1 - x^n$ for fixed x in $[0, 1]$, we obtain a differential equation for the generating function of N_t:

$$\frac{d}{dt}\mathbb{E}\ x^{N_t} = -(1-x)\ \mathbb{E}\ x^{N_t}\ R_t, \quad x \in [0,1], \quad t \in \mathbb{R}_+.$$

Also, taking $f(n) = n$ and $f(n) = n^2$ yield

$$\frac{d}{dt}\ \mathbb{E}\ N_t = \mathbb{E}\ R_t, \quad \frac{d}{dt}\ \mathbb{E}\ N_t^2 = \mathbb{E}\ R_t + 2\ \mathbb{E}\ R_t N_t.$$

Proof of Theorem 6.11

Let $g(\omega, s, z)$ be the integrand appearing on the right side of 6.12. The function g is the composition of the mappings $(\omega, s, z) \mapsto R_s(\omega, z)$ and $(r, z) \mapsto I(z, (0, r])$; the first mapping is measurable with respect to $\mathcal{F}^p \otimes \mathcal{B}_{\mathbb{R}_+}$ and $\mathcal{B}_{\mathbb{R}_+ \times \mathbb{R}_+}$ by the assumed $\mathcal{F}$-predictability of R; the second is obviously measurable with respect to $\mathcal{B}_{\mathbb{R}_+ \times \mathbb{R}_+}$ and $\mathcal{B}_{\mathbb{R}_+}$. Thus, g is in $\mathcal{F}^p \otimes \mathcal{B}_{\mathbb{R}_+}$.

Let F be a positive process in $\mathcal{F}^p$. Then, $G(\omega, t, z) = F_t(\omega) g(\omega, t, z)$ defines a positive process G in $\mathcal{F}^p \otimes \mathcal{B}_{\mathbb{R}_+}$. It follows from 6.12 and the definitions of g and G that

$$\int_{\mathbb{R}_+} F_t \, dN_t = \int_{\mathbb{R}_+ \times \mathbb{R}_+} M(dt, dz) \; G(t, z)$$

$$\int_{\mathbb{R}_+ \times \mathbb{R}_+} dt \; dz \; G(t, z) = \int_{\mathbb{R}_+} dt \; F_t \int_{\mathbb{R}_+} dz \; I(z, (0, R_t)) = \int_{\mathbb{R}_+} dt \; F_t R_t.$$

Taking expectations and using Theorem 6.2, we see that

6.16 $$\mathbb{E} \int_{\mathbb{R}_+} F_t \, dN_t = \mathbb{E} \int_{\mathbb{R}_+} dt \; F_t \; R_t,$$

which proves the claim that R is the $\mathcal{F}$-intensity of N.

There remains to prove 6.13 for f positive increasing on the integers. We start by observing that, since N is a counting process,

$$\begin{aligned} f(N_t) &= f(0) + \sum_{s \le t} [f(N_s) - f(N_{s-})] \\ &= f(0) + \int_{[0,t]} [f(N_{s-} + 1) - f(N_{s-})] dN_s. \end{aligned}$$

On the right side, the integrand is left-continuous in s and adapted and thus predictable, and it is positive since f is increasing. Thus, we may take the integrand to be $F_s \, 1_{[0,t]}(s)$ in 6.16, which yields

$$\begin{aligned} \mathbb{E} \; f(N_t) &= f(0) + \mathbb{E} \int_{[0,t]} ds \; [f(N_{s-} + 1) - f(N_{s-})] \; R_s \\ &= f(0) + \mathbb{E} \int_{[0,t]} ds \; [f(N_s + 1) - f(N_s)] \; R_s, \end{aligned}$$

where the last equality is justified by noting that the value of a Lebesgue integral does not change when the integrand is altered at countably many points (by replacing N_{s-} by N_s). □

In the setting of the preceding theorem, it follows from 6.12 that N depends on R and, hence, R depends on N. However, in non-mathematical terms, it is possible that "N depends on R but R is independent of N," as in the phrase "economy depends on the weather but the weather is independent of the economy." That is the situation if R is independent of M, as in

Exercise 2.36 for instance, and then N is conditionally Poisson given R. In the more interesting case where R depends on M and N, the process N appears implicitly on the right side of 6.12; hence, then, 6.12 is an integral equation to be solved for N. The term *self-exciting* is used in such cases where the past of N over $(0, t)$ defines, or affects, the value R_t.

Example: Branching with immigration

This is about the evolution of a system as follows. Primary particles arrive into the system according to a Poisson process with rate a; they form the original, $k = 0$, generation. The particles of the kth generation give births to the particles of the $(k+1)^{\text{th}}$ generation. Each particle, of whatever generation, gives births according to a Poisson process with rate b independent of the doings of all other particles. We are interested in the size N_t of the total population at time t. The following is the mathematical model for this story.

6.17 Model. Let M be a standard Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ (with mean Leb $\times$ Leb). Let $\mathcal{F}$ be the augmented filtration generated by M. Define N and R by setting $R_0 = N_0 = 0$ and

6.18 $$R_t = a + bN_{t-}, \qquad N_t = \int_{(0,t]\times\mathbb{R}_+} M(ds, dz)\; I(z, (0, R_s]), \quad t > 0,$$

where a and b are constants in $(0, \infty)$. Note that N is right-continuous and adapted to $\mathcal{F}$, and R is left-continuous and predictable. □

We describe the solution $N(\omega) : t \mapsto N_t(\omega)$ for a typical ω belonging to the almost sure event described in Theorem 2.18: the set D_ω of atoms of the measure M_ω is such that no two atoms arrive at the same time and there are only finitely many atoms in any bounded rectangle. We start with $N_0(\omega) = 0$; we look for the first t such that (t, z) is an atom in D_ω with size z under a; that t is the time $T_1(\omega)$ of the first jump for $N(\omega)$. We continue recursively: having picked $T_k(\omega)$, the time of the kth jump, we look for the first t after $T_k(\omega)$ such that (t, z) is in D_ω and z is under $a + kb$; that t is $T_{k+1}(\omega)$. In short, putting $T_0(\omega) = 0$ for convenience, we define

6.19 $$T_{k+1}(\omega) = \inf\{t > T_k(\omega) : \quad (t, z) \in D_\omega,\; z \le a + kb\}, \quad k \in \mathbb{N}.$$

Then, $T_1(\omega) < T_2(\omega) < \cdots$ are the successive jump times of $N(\omega)$, and

6.20 $$N_t(\omega) = k \iff T_k(\omega) \le t < T_{k+1}(\omega).$$

It follows from 6.19 and the Poisson nature of M that $T_1, T_2 - T_1, T_3 - T_2, \ldots$ are independent and exponentially distributed with respective parameters $a, a + b, a + 2b, \ldots$.

To show that N is a counting process, there remains to show that it does not explode in finite time, that is, $N_t < \infty$ almost surely for each t (which

implies that, for almost every ω, we have $N_t(\omega) < \infty$ for all $t < \infty$). This can be shown by showing that $\lim T_k = +\infty$ almost surely. It is easier and has more value to take a direct approach and show that N_t has finite expectation. We use 6.14 to this end:

$$\frac{d}{dt}\,\mathbb{E}\,N_t = \mathbb{E}\,R_t = \mathbb{E}\,(a + bN_{t-}) = a + b\,\mathbb{E}\,N_t$$

since $N_t = N_{t-}$ almost surely. Solving the differential equation with the initial condition $\mathbb{E}N_0 = 0$, we get

6.21 $$\mathbb{E}\,N_t = \frac{a}{b}(e^{bt} - 1), \quad t \in \mathbb{R}_+.$$

Finally, we consider the distribution of N_t for fixed t. To this end, we use 6.15 of Remark 6.14 with a well-chosen f to get a recursive formula for

6.22 $$p_k(t) = \mathbb{P}\{N_t = k\}, \quad k \in \mathbb{N}, \quad t \in \mathbb{R}_+.$$

Fix k. Let f be the indicator of the set $\{k+1, k+2, \ldots\}$. Then, $f(N_t)$ becomes the indicator of the event $\{N_t > k\}$, and $f(N_t + 1) - f(N_t)$ the indicator of $\{N_t = k\}$. And, on the event $\{N_t = k\}$, we have $R_t = a + bk$ almost surely since $N_t = N_{t-}$ almost surely. Thus 6.15 becomes

$$\frac{d}{dt}\,\mathbb{P}\{N_t > k\} = (a + bk)\mathbb{P}\{N_t = k\}, \quad k \in \mathbb{N}.$$

Equivalently,

$$\begin{aligned} \frac{d}{dt}p_0(t) &= -ap_0(t), \\ \frac{d}{dt}p_k(t) &= -(a + bk)p_k(t) + (a + bk - b)p_{k-1}(t), \quad k \geq 1, \end{aligned}$$

with the obvious initial conditions stemming from $p_0(0) = 1$. This system can be solved recursively:

6.23 $$p_0(t) = e^{-at}, \quad p_k(t) = (a + kb - b)\int_0^t e^{-(a+bk)(t-s)}p_{k-1}(s)\,ds, \quad k \geq 1.$$

Example: Self-exciting shot processes

These are similar to processes described in the preceding example, except that the immigration rate is ae^{-ct} at time t and that each particle in the system gives births at rate be^{-cu} when the particle's age is u. The preceding example is the particular case where $c = 0$. We are interested in the number N_t of particles in the system at time t.

6.24 MODEL. Let M be a standard Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$. Let $\mathcal{F}$ be the augmented filtration generated by it. Define N and R by setting $N_0 = 0$ and $R_0 = a$ and

6.25 $$R_t = ae^{-ct} + \int_{(0,t)} be^{-c(t-s)}\, dN_s, \quad t > 0,$$

6.26 $$N_t = \int_{[0,t]\times\mathbb{R}_+} M(ds, dz)\; I(z, (0, R_s]), \quad t > 0,$$

where a, b, c are constants in $(0, \infty)$. Note that N is right-continuous, R is left-continuous, and both are adapted to $\mathcal{F}$. □

The process R of 6.25 is a shot-noise process driven by N; it resembles that in Example 2.12, but now N is driven by R in turn. It is clear from 6.26 that N increases by jumps of size one. To show that it is a counting process, there remains to show that $N_t < \infty$ almost surely. This last point is easy: note that R and N of Model 6.24 are dominated by the respective ones of the Model 6.17; hence, $\mathbb{E}\, N_t \le \frac{a}{b}e^{bt}$ by 6.21; in view of Remark 6.14, the exact value can be obtained by integrating 6.35 below.

The solution to the coupled system 6.25–6.26 is as follows for the typical "good" ω of Theorem 2.18: The first jump of $N(\omega)$ occurs at the first time t where (t, z) is an atom of M_ω with $z \le ae^{-bt}$; for u in $(0, t]$ we have $R_u(\omega) = ae^{-cu}$, and $N_t(\omega) = 1$ obviously. Assuming that s is the kth jump time and $R_s(\omega) = r$ and $N_s(\omega) = k$, the time of next jump is the smallest t in (s, ∞) where (t, z) is an atom of M_ω having $z \le (r + b)e^{-c(t-s)}$ and, then, $R_u(\omega) = (r + b)e^{-c(u-s)}$ for all u in $(s, t]$, and $N_u(\omega) = k$ for u in $[s, t)$, and $N_t(\omega) = k + 1$.

6.27 *Markov property.* The process R is a Markov process, that is, for every time t, the future process $\hat{R} = \{R_{t+u}; u \in \mathbb{R}_+\}$ is conditionally independent of the past $\mathcal{F}_t$ given the present state R_t. Moreover, given that $R_t = x$, the conditional law of $\hat{R}$ is the same as the law of R starting from $a = x$.

To see this we re-write 6.25 and 6.26 for the time $t + u$. Define $\hat{M}(A) = M(\hat{A})$ with $\hat{A} = \{(t + u, z) : (u, z) \in A\}$, and observe that $\hat{M}$ is independent of $\mathcal{F}_t$ and has the same law as M, that is, Poisson on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $\text{Leb} \times \text{Leb}$. Define

6.28 $$\hat{R}_u = R_{t+u}, \quad \hat{N}_u = N_{t+u} - N_t, \quad u \in \mathbb{R}_+.$$

It follows from 6.25 and 6.26 after some re-arrangement that

6.29 $$\hat{R}_u = \hat{R}_0 e^{-cu} + \int_{(0,u)} be^{-c(u-s)} d\hat{N}_s$$

6.30 $$\hat{N}_u = \int_{[0,u]\times\mathbb{R}_+} \hat{M}(ds, dz)\; I(z, (0, \hat{R}_s]).$$

These show that the pair $(\hat{N}, \hat{R})$ satisfies the equations 6.25 and 6.26 with $a = \hat{R}_0$ and M replaced with $\hat{M}$. Since the solution to 6.25–6.26 is unique, and since $\hat{M}$ is standard Poisson independent of $\mathcal{F}_t$, we conclude that $(\hat{N}, \hat{R})$ is conditionally independent of $\mathcal{F}_t$ given $\hat{R}_0 = R_t$ and $\hat{N}_0 = 0$.

Indeed we have shown more than what was listed in 6.27: The pair (N, R) is a Markov process as well as the process R. □

6.31 *Distribution of* R_t. Let f be a bounded function that is differentiable and assume that its derivative f' is also bounded. Define

6.32 $$Gf(x) = -cxf'(x) + x[f(x+b) - f(x)].$$

Then,

6.33 $$\mathbb{E}\ f(R_t) = f(a) + \int_{(0,t)} ds\ \mathbb{E}\ Gf(R_s).$$

To prove this, we start by noting that each jump of R is of size b and that, between the jumps, R decays exponentially at rate c. It follows that

$$f(R_t) = f(R_0) - \int_{(0,t)} cR_s f'(R_s)ds + \int_{(0,t)} [f(R_s + b) - f(R_s)]dN_s.$$

Within the last integral, the integrand is predictable, because it is a continuous function of a left-continuous adapted process. Thus, by the meaning of intensity (see 6.6)

$$\mathbb{E}\ f(R_t) = f(a) - \mathbb{E}\ \int_{(0,t)} cR_s f'(R_s)ds + \mathbb{E}\ \int_{(0,t)} R_s[f(R_s+b) - f(R_s)]ds;$$

here, the boundedness of f and f' is used to ensure that the expectations are well-defined. This last formula is exactly the result that was to be shown. □

In practice, it is generally easier to use 6.33 in its differential form:

6.34 $$\frac{d}{dt}\ \mathbb{E}\ f(R_t) = \mathbb{E}\ Gf(R_t), \quad \mathbb{E}\ f(R_0) = f(a).$$

Here are several quick uses: Taking $f(x) = x$ and solving 6.34,

6.35 $$\mathbb{E}\ R_t = ae^{-(c-b)t}, \quad t \in \mathbb{R}_+.$$

Taking $f(x) = e^{-px}$ and noting that

$$Gf(x) = (1 - e^{-pb} + cp)\frac{\partial}{\partial p}e^{-px},$$

we see that

6.36 $$u(t,p) = \mathbb{E}\ e^{-pR_t}, \quad t, p \in \mathbb{R}_+,$$

is the solution to

6.37
$$\frac{\partial}{\partial t}u(t,p) = (1 - e^{-pb} + cp)\frac{\partial}{\partial p}u(t,p)$$

with boundary conditions $u(t,0) = 1$ and $u(0,p) = e^{-ap}$. □

Going back to 6.35, we observe that $\mathbb{E}\, R_t = a$ for all t if $c = b$. Otherwise, if $c > b$, the expectation goes to 0 as $t \to \infty$, and, if $c < b$, it goes to $+\infty$. Indeed, these observations can be sharpened:

6.38 *Martingales.* The process R is a martingale if $c = b$, supermartingale if $c > b$, and submartingale if $c < b$. If $c \geq b$,

$$R_\infty = \lim_{t\to\infty} R_t$$

exists and is an integrable random variable with $0 \leq \mathbb{E}\, R_\infty \leq a$. In particular, when $c > b$, we have $R_\infty = 0$ almost surely.

To show these claims, we go back to the Markov property 6.27 for R to conclude that the conditional expectation of R_{t+u} given $\mathcal{F}_t$ is equal to $\mathbb{E}R_u$ with a replaced by R_t, that is, in view of 6.35,

6.39
$$\mathbb{E}_t\ R_{t+u} = R_t\ e^{-(c-b)u}, \quad t, u \in \mathbb{R}_+.$$

This shows that R is a martingale if $c = b$, supermartingale if $c > b$, and submartingale if $c < b$. In the first two cases, that is, if $c \geq b$, we have a positive supermartingale, which necessarily converges almost surely to some integrable random variable $R_\infty \geq 0$; that is by the convergence theorems for such. Moreover, by Fatou's lemma,

$$\mathbb{E}\ R_\infty = \mathbb{E}\ \liminf R_t \leq \liminf\ \mathbb{E}\ R_t,$$

which shows that $0 \leq \mathbb{E}R_\infty \leq a$ via 6.35. In particular, this becomes $\mathbb{E}R_\infty = 0$ when $c > b$, which implies that $R_\infty = 0$.

Compensators

Not every counting process has an intensity process. This is to compensate for this lack and introduce the general ideas involved.

Let N be a counting process adapted to some filtration $\mathcal{F}$ and suppose that $\mathbb{E}N_t < \infty$ for every t in $\mathbb{R}_+$. Then, it is known that there exists an increasing $\mathcal{F}$-predictable process $C = (C_t)$ such that

6.40
$$\tilde{N}_t = N_t - C_t, \quad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale; this follows from the continuous time version of Doob's decomposition applied to the submartingale N. The process C is called the *compensator* for N relative to $\mathcal{F}$, or *dual-predictable projection* of N. The

term *compensator* may be justified by noting that $\tilde{N}$ is a martingale if and only if

6.41 $$\mathbb{E}\int_{\mathbb{R}_+} F_t\, dN_t = \mathbb{E}\int_{\mathbb{R}_+} F\, dC_t$$

for every positive (or bounded) $\mathcal{F}$-predictable process F.

Comparing 6.41 with Definition 6.6 of intensities, we conclude the following: An $\mathcal{F}$-predictable process R is the intensity of N relative to $\mathcal{F}$ if and only if

6.42 $$C_t = \int_0^t ds\; R_s, \quad t \in \mathbb{R}_+.$$

is the compensator of N relative to $\mathcal{F}$. Thus, existence of an intensity has to do with the absolute continuity of C.

The following is the counterpart to the construction of Theorem 6.11. It is analogous to Theorem 6.18 of Chapter V on non-stationary Poisson processes. When time is reckoned with the clock C, the process N appears to be Poisson with unit rate.

6.43 Theorem. *Let N be a counting process adapted to $\mathcal{F}$ and with $\mathbb{E}\, N_t < \infty$ for all t. Let C be its compensator relative to $\mathcal{F}$. Suppose that C is continuous and* $\lim_{t\to\infty} C_t = +\infty$. *Define, for u in $\mathbb{R}_+$,*

6.44 $$S_u = \inf\{t : C_t > u\}, \quad \hat{N}_u = N_{S_u}, \quad \hat{\mathcal{F}}_u = \mathcal{F}_{S_u}.$$

Then, $\hat{N}$ is a Poisson process with unit rate with respect to $\hat{\mathcal{F}}$, and for each t, almost surely,

6.45 $$N_t = \hat{N}_{C_t}.$$

Remark. We regard C_t as time shown on a rigged clock when the standard time is t. Then, S_u becomes the standard time when the clock shows u. If N is an arrival process with standard time parameter, then $\hat{N}$ is the same arrival process in clock time. In general C and $\hat{N}$ are dependent. In the special case that C and $\hat{N}$ are independent, the formula 6.45 is very convenient: for instance, then, the conditional probability that $N_t - N_s = k$ given $\mathcal{F}_s$ is equal to, for $0 \le s < t$ arbitrary,

$$\mathbb{E}_s e^{-(C_t - C_s)}(C_t - C_s)^k/k!, \quad k \in \mathbb{N}.$$

In other words, if $\hat{N}$ and C are independent, then the conditional law of N given C is that of a non-stationary Poisson process with mean function C. The further special case where C is deterministic yields a non-stationary Poisson process N.

Proof. Let S and T be stopping times with $S \le T < \infty$. Let $H \in \mathcal{F}_S$ and define $F_t = 1_H\; 1_{(S,T]}(t)$. Then, F is predictable, and using 6.41 with this F shows that

6.46 $$\mathbb{E}_S(N_T - N_S) = \mathbb{E}_S(C_T - C_S)\,.$$

Fix $0 \le u < v$ and let $S = S_u$ and $T = S_v$; these are finite stopping times in view of 6.44 and the assumption that $\lim C_t = \infty$ while $C_t < \infty$ almost surely for every t in $\mathbb{R}_+$ (since $\mathbb{E}C_t = \mathbb{E}N_t < \infty$). Now, $N_T - N_S = \hat{N}_v - \hat{N}_u$ by the definition of $\hat{N}$, and $C_T - C_S = v - u$ by 6.44 and the continuity of C. It now follows from 6.46 that $\hat{N}_v - \hat{N}_u$ is independent of $\hat{\mathfrak{F}}_u$ and has mean equal to $v-u$. It now follows from the characterization theorem 5.5 (or Theorem 6.18 of Chapter V) that $\hat{N}$ is Poisson with respect of $\hat{\mathfrak{F}}$ with unit rate.

There remains to show that, for each t, 6.45 holds almost surely. If C is strictly increasing, then $S_{C_t} = t$ and the equality 6.45 is without question. If C is not such, then $S_{C_t} \ge t$ and the strict inequality may hold for some outcomes ω. For such ω, however, $C(\omega)$ remains flat and equal to $C_t(\omega)$ over the interval $(t, S_{C_t(\omega)}(\omega)]$. In view of 6.46, almost surely on the set of all such ω, the process N can have no jumps over the interval $[t, S_{C_t}]$. Thus, $N_t = \hat{N}_{C_t}$ almost surely. □

Exercises

6.47 *Shot process.* For the process R of model 6.24, show that

$$\text{Var } R_t = ab^2 \int_t^{2t} e^{-(c-b)s} ds.$$

6.48 *Continuation.* Show that, for f as in 6.31 and $g : \mathbb{N} \mapsto \mathbb{R}$ arbitrary bounded,

$$\frac{d}{dt}\mathbb{E}\ f(R_t)g(N_t) = \mathbb{E}\ h(R_t, N_t)$$

where

$$h(x,n) = -cxf'(x)g(n) + x[f(x+a)g(n+1) - f(x)g(n)].$$

Use this to derive a partial differential equation for

$$u(t,p,q) = \mathbb{E}\ \exp_-\ (pR_t + qN_t).$$

6.49 *Covariance density for N of 6.24.* We switch to regarding N as a random measure on $\mathbb{R}_+$. Notationally, dN_t becomes $N(dt)$, and $d\hat{N}_u$ becomes $N(t + du)$ when $\hat{N}_u = N_{t+u} - N_t$. Notice that 6.8 and 6.39 may be presented as

$$\mathbb{E}_t N(t+du) = du\ \mathbb{E}_t R_{t+u} = du\ R_t\ e^{-(c-b)u}.$$

Show that, for $t, u > 0$,

$$\frac{1}{dt\ du}\left[\mathbb{E}\ N(dt)N(t+du) - \mathbb{E}\ N(dt)\ \mathbb{E}\ N(t+du)\right] = e^{-(c-b)u}\ \text{Var } R_t.$$

The left side is called the covariance density at $(t, t+u)$.

6.50 *Branching formulation.* This is another formulation of the model 6.24. We view the population as the sum of generations $0, 1, \ldots$; The original generation consists of particles that arrived into the system according to a Poisson process N_0 with deterministic intensity process $R_0(t) = ae^{-ct}$, $t \in \mathbb{R}_+$. Having described the generation k, we recall that each particle of generation k, if it starts life at time s, gives births to generation $k+1$ particles at the rate be^{-cu} at the time $s+u$. Thus, the births of $(k+1)^{\text{th}}$ generation particles form a conditionally Poisson process N_{k+1} given the intensity process

$$R_{k+1}(t) = \int_{(0,t)} N_k(ds)\ be^{-c(t-s)}.$$

The process N is the sum of all generations,

$$N(A) = \sum_{k=0}^{\infty} N_k(A), \quad \text{Borel } A \subset \mathbb{R}_+.$$

The advantage of this formulation is that we have a chain $N_0, N_1, N_2, \ldots$ of random counting measures on $\mathbb{R}_+$, each one determines the intensity process for the next, each one being conditionally Poisson given its own intensity.

6.51 *Hawkes processes.* These are processes much like N of the model 6.24, but with some slight generality: N is still defined by 6.26, but 6.25 is replaced with

$$R_t = a + \int_{(0,t)} g(t-s)\ dN_s, \quad t > 0,$$

where g is some deterministic positive Borel function.

a) Suppose that g is bounded. Show that, then, $\mathbb{E}\ N_t < \infty$ for all t and there is a unique counting process N with this intensity process. Hint: Compare this with the model 6.17.

b) Compute $\mathbb{E}\ R_t$ and $\mathbb{E}\ N_t$ exactly.

c) Discuss the limiting behavior of R_t as $t \to \infty$ in the case when g is Lebesgue-integrable.

6.52 *Continuation.* Consider the processes R and N of the preceding exercise for the special case

$$g(u) = 1_{[b,c)}(u), \quad u \in \mathbb{R}_+,$$

where $0 < b < c$ are constants. This corresponds to the case each particle gives births to new ones at unit rate starting when it is at age b and ending when it dies at age c. Describe the solution N. Compute $\mathbb{E}\ N_t$ explicitly.

6.53 *Departures from an $M/M/\infty$ queue.* Let L be a Poisson process with rate a. Let M be as in Theorem 6.11. Suppose that L and M are independent and let $\mathcal{F}$ be the augmented filtration generated by them. Define N by 6.12 with

$$R_t = b \cdot (L_{t-} - N_{t-})^+, \quad t > 0,$$

with $R_0 = N_0 = 0$. Note that $R_t = 0$ when $L_{t-} \leq N_{t-}$, which implies that $L_t \geq N_t$ for all t.

a) Show that

$$\mathbb{P}\{L_{t+u} - L_t = N_{t+u} - N_t = 0 | L_t - N_t = k\} = e^{-(a+bk)u}.$$

b) Show that $Q = L - N$ can be regarded as the queue size process in Exercise 3.35 with the further assumption that each service lasts an exponentially distributed time with parameter b.

Chapter VII

Lévy Processes

This chapter is on Lévy processes with state space $\mathbb{R}^d$, their structure and general properties. Section 1 introduces them and gives a constructive survey of the range of behaviors. Section 2 illustrates those constructions in the case of stable processes, a special class.

Section 3 re-introduces Lévy processes in a modern setting, discusses the Markov and strong Markov properties for them, and shows the special nature of the filtrations they generate. Section 4 characterizes the three basic processes, Poisson, compound Poisson, and Wiener, in terms of the qualitative properties of the sample paths. Section 5 is on the famous characterization theorem of Itô and Lévy, showing that every Lévy process has the form constructed in Section 1; we follow Itô's purely stochastic treatment.

Section 6 is on the use of increasing Lévy processes in random time changes, an operation called subordination with many applications. Finally, in Section 7, we describe some basic results on increasing Lévy processes; these are aimed at applications to theories of regeneration and Markov processes.

The special case of Wiener processes is left to the next chapter for a deeper treatment.

1 Introduction

Let $(\Omega, \mathcal{H}, \mathbb{P})$be a probability space. Let $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration on it. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a stochastic process with state space $\mathbb{R}^d$; here, $d \geq 1$ is the dimension, and the relevant σ-algebra on $\mathbb{R}^d$ is the Borel one.

1.1 Definition. *The process X is called a* Lévy process in $\mathbb{R}^d$ *with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$ and*

a) for almost every ω, the path $t \mapsto X_t(\omega)$ is right-continuous and left-limited starting from $X_0(\omega) = 0$, and

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_7,

b) for every t and u in $\mathbb{R}_+$, the increment $X_{t+u} - X_t$ is independent of $\mathcal{F}_t$ and has the same distribution as X_u.

Let $\mathcal{G} = (\mathcal{G}_t)_{t\in\mathbb{R}_+}$ be the filtration generated by X. If X is a Lévy process with respect to $\mathcal{F}$, then it is such with respect to $\mathcal{G}$ automatically. It is called a Lévy process, without mentioning a filtration, if it is such with respect to $\mathcal{G}$.

In the preceding definition, the first condition is on the regularity of paths. The second condition implies that X has *stationary and independent increments*: $X_{t+u} - X_t$ has the same distribution for all t, and the increments $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for all choices of $n \geq 2$ and times $0 \leq t_0 < t_1 < \cdots < t_n$. Conversely, if X has stationary and independent increments, then it fulfills the condition 1.1b with $\mathcal{F} = \mathcal{G}$.

Every constant multiple of a Lévy process in $\mathbb{R}^d$ is again Lévy. The sum of a finite number of independent Lévy processes in $\mathbb{R}^d$ is again Lévy. Given a Lévy process X in $\mathbb{R}^d$ and a $d' \times d$ matrix c, the process cX is a Lévy process in $\mathbb{R}^{d'}$; in particular, every linear combination of the components of X is a Lévy process in $\mathbb{R}$; every component of X is a Lévy process in $\mathbb{R}$ - the components generally depend on each other.

1.2 Example. The simplest (and trivial) Lévy process in $\mathbb{R}^d$ is the pure-drift: it has the form $X_t = bt$ where b is a fixed vector in $\mathbb{R}^d$. Next, we recall the definitions of some Lévy processes introduced in earlier chapters.

a) According to Definition V.2.15, a Wiener process W is a Lévy process in $\mathbb{R}$ that has continuous paths and has the Gaussian distribution with mean 0 and variance u for its increments $W_{t+u} - W_t$. It is the basic continuous Lévy process: The most general continuous Lévy process in $\mathbb{R}$ has the form

$$X_t = bt + cW_t, \quad t \in \mathbb{R}_+,$$

where b and c are constants in $\mathbb{R}$. A similar result holds for processes in $\mathbb{R}^d$, in which case b is a vector in $\mathbb{R}^d$, and c is a $d \times d'$ matrix, and W is a d'-dimensional Wiener process (whose components are independent Wiener processes). See Theorem 4.3.

b) *Poisson processes.* The initial definition was given in Definition V.2.20: a Poisson process N with rate c is a Lévy process that is a counting process having the Poisson distribution with mean cu for its increments $N_{t+u} - N_t$. A list of characterizations were given in Section 5 of the preceding chapter, and also a martingale characterization in Theorem V.6.13. We shall add one more in Section 4: a Lévy process whose increments are Poisson distributed is necessarily a counting process (and, hence, is a Poisson process).

c) *Compound Poisson process.* These were introduced in Section 3 of the preceding chapter as follows. Let N be a Poisson process. Independent of it, let (Y_n) be an independency of identically distributed $\mathbb{R}^d$-valued random variables. Define

$$X_t = \sum_{n=1}^{\infty} Y_n 1_{\{n \leq N_t\}}, \quad t \in \mathbb{R}_+.$$

Then, X is a Lévy process in $\mathbb{R}^d$. Its every path is a step function; its jumps occur at the jump times of N, and the sizes of successive jumps are $Y_1, Y_2, \ldots$. We shall show in Theorem 4.6 that, conversely, every Lévy process whose paths are step functions is a compound Poisson process.

d) *Increasing Lévy processes.* According to Definition VI.4.5, these are Lévy processes in $\mathbb{R}$ whose paths are increasing. Equivalently, they are Lévy processes with state space $\mathbb{R}_+$, because the positivity of X_u and the stationarity of $X_{t+u} - X_t$ imply that every increment is positive. Every Poisson process is an increasing Lévy process. So is every compound Poisson process with positive jumps (with $\mathbb{R}_+$-valued Y_n in the preceding remark). So are gamma processes, so are stable processes with indices in $(0, 1)$; see 4.9, 4.10, 4.19, 4.20 of Chapter VI for these, and also Propositions 4.6 and 4.14 there for general constructions. It will become clear that every increasing Lévy process has the form given in Proposition VI.4.6; see Remark 5.4b to come.

Infinite divisibility, characteristic exponent

Recall that a random variable is said to be infinitely divisible if, for every integer n, it can be written as the sum of n independent and identically distributed random variables. Let X be a Lévy process in $\mathbb{R}^d$. For $t > 0$ fixed and $n \geq 1$, lettting $\delta = t/n$, we can write X_t as the sum of the increments over the intervals $(0, \delta], (\delta, 2\delta], \cdots, (n\delta - \delta, n\delta]$, and those increments are independent and identically distributed. Thus, X_t is infinitely divisible for every t, and so is every increment $X_{t+u} - X_t$. It follows that the characteristic function of X has the form

1.3 $$\mathbb{E}\, e^{ir \cdot X_t} = e^{t\psi(r)}, \qquad t \in \mathbb{R}_+,\ r \in \mathbb{R}^d\ ;$$

here, on the left, $r \cdot x = r_1 x_1 + \cdots + r_d x_d$, the inner product of r and x in $\mathbb{R}^d$. On the right side, ψ is some complex-valued function having a specific form; it is called the characteristic exponent of X. Its form is given by the Lévy-Khinchine formula; see 1.31 and 1.33 below. Its derivation is basically a corollary to Itô-Lévy decomposition of Theorem 5.2 to come.

Means and variances

Let X be a Lévy process in $\mathbb{R}^d$. It is possible that $\mathbb{E}X_t$ does not exist; this is the case, for instance, if X is a compound Poisson process as in Example 1.2c and the Y_n do not have expected values. Or, it is possible that $\mathbb{E}X_t$ is well-defined but is equal to infinity in some components. However, if the means and variances of the components of the random vector X_t are well-defined, then they must be linear in t, that is,

1.4 $$\mathbb{E}\ X_t = at, \quad \mathrm{Var} X_t = vt, \quad t \in \mathbb{R}_+.$$

This is a consequence of the stationarity and independence of the increments; a is a fixed vector in $\mathbb{R}^d$, and $\mathrm{Var} X_t$ is notation for the covariance matrix of

X_t, and v is a fixed symmetric $d \times d$ matrix that is positive definite, that is, $v_{ij} = v_{ji}$ for all i and j, and $r \cdot vr \geq 0$ for every r in $\mathbb{R}^d$.

Continuity in distribution

Consider 1.3 and note its continuity in t. Recall that the convergence in distribution is equivalent to the convergence of the corresponding characteristic functions; see Corollary III.5.19. Thus, (X_{t_n}) converges in distribution to X_t for every sequence (t_n) with limit t. The following is the same statement using the definition of convergence in distribution.

1.5 PROPOSITION. *Suppose that X is a Lévy process in $\mathbb{R}^d$. Then, $t \mapsto \mathbb{E}\, f \circ X_t$ is continuous for every bounded continuous function $f : \mathbb{R}^d \mapsto \mathbb{R}$.*

Probability law of X

Suppose that X is Lévy. Then, its probability law is determined by the distribution π_t of X_t for any one $t > 0$, or equivalently, by the characteristic exponent ψ appearing in 1.3. To see this, first note that the Fourier transform of π_t is $e^{t\psi}$; if it is known for one t, then it is known for all t.

Next, consider the finite-dimensional distributions of X: consider the distribution of $(X_s, X_t, \ldots, X_u, X_v)$ for finitely many times, $0 < s < t < \cdots < u < v$. That distribution is determined by the distribution of $(X_s, X_t - X_s, \ldots, X_v - X_u)$, and the latter is the product measure $\pi_s \times \pi_{t-s} \times \cdots \times \pi_{v-u}$ in view of the independence and stationarity of the increments.

Regularity of the paths and jumps

Suppose that X is a Lévy process in $\mathbb{R}^d$. Fix an outcome ω for which the regularity properties 1.1a hold. This means that the limits

1.6 $$X_{t-}(\omega) = \lim_{s \uparrow t} X_s(\omega), \quad X_{t+}(\omega) = \lim_{u \downarrow t} X_u(\omega)$$

exist for every t in $\mathbb{R}_+$ (with the convention that $X_{t-}(\omega) = 0$ for $t = 0$), the limits belong to $\mathbb{R}^d$, and $X_{t+}(\omega) = X_t(\omega)$ by right-continuity. If the two limits differ, then we say that the path $X(\omega)$ jumps from its left-limit $X_{t-}(\omega)$ to its right-hand value $X_t(\omega) = X_{t+}(\omega)$. The difference

1.7 $$\Delta X_t(\omega) = X_t(\omega) - X_{t-}(\omega),$$

if non-zero, is called the *size* of the jump at time t and its length $|\Delta X_t(\omega)|$ is called the *jump magnitude*. The path $X(\omega)$ can have no discontinuities other than the jump-type described.

Let D_ω be the discontinuity set for the path $X(\omega)$, that is,

1.8 $$D_\omega = \{t > 0 \, : \, \Delta X_t(\omega) \neq 0\}.$$

If X is continuous, then D_ω is empty for almost every ω. If X is Poisson or compound Poisson plus some continuous process, then, for almost every ω, the set D_ω is an infinite countable set, but $D_\omega \cap (s,u)$ is finite for all $0 \le s < u < \infty$. For all other processes X, for almost every ω, the set D_ω is still infinite but with the further property that $D_\omega \cap (s,u)$ is infinite for all $0 \le s < u < \infty$. This last property is apparent for gamma and stable processes of Example VI.4.9 and VI.4.10, and will follow from Itô-Lévy decomposition in general; see Theorem 5.2.

1.9 REMARK. However, for every $\varepsilon > 0$, there can be at most finitely many t in $D_\omega \cap (s,u)$ for which the magnitude $|\Delta X_t(\omega)|$ exceeds ε. For, otherwise, if there were infinitely many such jump times for some $\varepsilon > 0$, then Bolzano-Weierstrass theorem would imply that there must exist a sequence (t_n) of such times that converges to some point t in $[s,u]$, and then at least one of the limits 1.6 must fail to exist.

Pure-jump processes

These are processes in $\mathbb{R}^d$ where X_t is equal to the sum of the sizes of its jumps during $[0,t]$; more precisely, for almost every ω,

1.10
$$X_t(\omega) = \sum_{s \in D_\omega \cap [0,t]} \Delta X_s(\omega), \quad t \in \mathbb{R}_+,$$

where the sum on the right side converges absolutely, that is, where

1.11
$$V_t(\omega) = \sum_{s \in D_\omega \cap [0,t]} |\Delta X_s(\omega)| < \infty.$$

Indeed, every such process has bounded variation over bounded intervals, and $V_t(\omega)$ is the total variation of the path $X(\omega)$ over $[0,t]$.

Every increasing Lévy process without drift is a pure-jump Lévy process, so is the difference of two such independent processes. The following constructs such processes in general. We shall see later that every pure-jump Lévy process in $\mathbb{R}^d$ has the form given in this theorem.

1.12 THEOREM. *Let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean measure $\text{Leb} \times \lambda$, where the measure λ on $\mathbb{R}^d$ has $\lambda\{0\} = 0$ and*

1.13
$$\int_{\mathbb{R}^d} \lambda(dx)(|x| \wedge 1) < \infty.$$

Then, for almost every ω, the integral

1.14
$$X_t(\omega) = \int_{[0,t]\times\mathbb{R}^d} M_\omega(ds,dx)x$$

converges absolutely for every t, and the path $X(\omega)$ has bounded variation over $[0,t]$ for every t in $\mathbb{R}_+$. The process X is a pure-jump Lévy process in $\mathbb{R}^d$, and its characteristic exponent is

1.15 $$\psi(r) = \int_{\mathbb{R}^d} \lambda(dx)(e^{ir\cdot x} - 1), \quad r \in \mathbb{R}^d.$$

1.16 REMARK. *Lévy measure.* The measure λ determines the probability laws of M and X. It is called the Lévy measure of X. It regulates the jumps: for every Borel subset A of $\mathbb{R}^d$ with $\lambda(A) < \infty$, the jump times of X with corresponding sizes belonging to A form the counting process $t \mapsto M((0,t] \times A)$, and the latter is a Poisson process with rate $\lambda(A)$. The condition that $\lambda\{0\} = 0$ is for reasons of convenience: to prevent linguistic faults like "jumps of size 0," and also to ensure that $X(\omega)$ and M_ω determine each other uniquely for almost every ω. The condition 1.13 is essential. It is satisfied by every finite measure. More interesting are infinite measures that satisfy it; to such measures there correspond pure-jump processes that have infinitely many jumps during every interval (s,t) with $s < t$; but, of those jumps, only finitely many may exceed ε in magnitude however small $\varepsilon > 0$ may be; see Remark 1.9.

Proof. Let $\hat{M}$ be the image of M under the mapping $(s,x) \mapsto (s,|x|)$ from $\mathbb{R}_+ \times \mathbb{R}^d$ into $\mathbb{R}_+ \times \mathbb{R}_+$, and $\hat{\lambda}$ the image of λ under the mapping $x \mapsto |x|$. Then, $\hat{M}$ is Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $Leb \times \hat{\lambda}$. Note that $\hat{\lambda}\{0\} = 0$ and 1.13 is equivalent to

1.17 $$\int_{\mathbb{R}_+} \hat{\lambda}(dv)(v \wedge 1) < \infty,$$

which, in particular, implies that $\hat{\lambda}(\varepsilon, \infty) < \infty$ for every $\varepsilon > 0$.

Thus, by Proposition VI.2.18, we can select an almost sure event Ω' such that, for every ω in it, the measure $\hat{M}_\omega$ is a counting measure, has no atoms in $\{0\} \times \mathbb{R}_+$ and no atoms in $\mathbb{R}_+ \times \{0\}$, and has at most one atom in $\{t\} \times \mathbb{R}_+$ no matter what t is.

On the other hand, for each time t,

1.18 $$V_t = \int_{[0,t]\times\mathbb{R}_+} \hat{M}(ds,dv)v = \int_{[0,t]\times\mathbb{R}^d} M(ds,dx)|x|$$

is positive and real-valued almost surely in view of 1.17 and Proposition VI.2.13. Let Ω_t be the almost sure event involved, and define Ω'' to be the intersection of Ω_t over t in $\mathbb{N}$.

Fix an outcome ω in the almost sure event $\Omega' \cap \Omega''$. The mapping $t \mapsto V_t(\omega)$ from $\mathbb{R}_+$ into $\mathbb{R}_+$ is right-continuous and increasing starting from the

origin; and it has a jump of size v at time s if and only if M_ω has an atom (s,x) with $|x| = v$. It follows that the integral in 1.14 converges absolutely for all times t, and we have

$$\sum_{s\le t} |\Delta X_s(\omega)| = V_t(\omega), \quad \sum_{s\le t} \Delta X_s(\omega) = X_t(\omega).$$

Hence, X is of the pure-jump type and is right-continuous and left-limited starting from the origin, and its total variation over $[0,t]$ is equal to V_t.

It is immediate from 1.14 and the Poisson character of M that X has stationary and independent increments. The form 1.15 for the characteristic exponent follows from 1.3, 1.14, and Theorem VI.2.9. □

1.19 Remark. *Total variation.* The preceding proof has shown, in addition, that the total variation process V is defined by 1.18 as well, and that it is a pure-jump increasing Lévy process. Its Lévy measure is the image of λ under the mapping $x \mapsto |x|$. The path $X(\omega)$ has a jump of some size x at time t if and only if $V(\omega)$ has a jump of size $|x|$ at the same time t.

1.20 Remark. *Poisson and compound Poisson.* If the dimension $d = 1$, and $\lambda = c\delta_1$ (recall that δ_x is Dirac at x), then X of the last theorem becomes a Poisson process with rate c. For arbitrary d, if λ is a finite measure on $\mathbb{R}^d$, then 1.13 holds automatically and X is a compound Poisson process as in Example 1.2c: its jump times form a Poisson process N with rate $c = \lambda(\mathbb{R}^d)$, and the sizes Y_n of its jumps are independent of N and of each other and have the distribution $\mu = \frac{1}{c}\lambda$ on $\mathbb{R}^d$. Its total variation process V is an increasing compound Poisson process in $\mathbb{R}_+$; the jump times of V form the same Poisson process N, but the jump sizes are the $|Y_n|$.

1.21 Example. *Gamma, two-sided and symmetric.* Recall Example VI.4.6, the gamma process with shape rate a and scale parameter c. It is an increasing pure-jump Lévy process in $\mathbb{R}_+$. Its Lévy measure has the density ae^{-cx}/x for x in $(0,\infty)$ and puts no mass elsewhere. Its value at t has the gamma distribution with shape index at and scale c.

Let X^+ and X^- be independent gamma processes. Then,

$$X = X^+ - X^-$$

is a pure-jump Lévy process in $\mathbb{R}$; the distribution of X_t is not gamma; nevertheless, X may be called a two-sided gamma process; see Exercises 1.47 and 1.48 for some observations. In the special case where X^+ and X^- have the same law, that is, if they have the same shape rate a and the same scale parameter c, then the Lévy measure of X is given by

$$\lambda(dx) = dx\, a\frac{e^{-c|x|}}{|x|}, \quad x \in \mathbb{R}\backslash\{0\},$$

with $\lambda\{0\} = 0$; in this case, we call X a symmetric gamma process with shape rate a and scale parameter c. The distribution of X_t is not gamma and cannot be expressed explicitly; however, the characteristic function is

$$\mathbb{E}\, e^{irX_t} = \left(\frac{c}{c-ir}\right)^{at}\left(\frac{c}{c+ir}\right)^{at} = \left(\frac{c^2}{c^2+r^2}\right)^{at}, \quad r \in \mathbb{R}.$$

The total variation process $V = X^+ + X^-$ is a gamma process with shape rate $2a$ and scale parameter c. See Exercise 1.48 and also 6.26 for d-dimensional analogs of X.

Compensated sums of jumps

This is to introduce Lévy processes driven by Poisson random measures as above, but whose paths may have infinite total variation over every time interval of strictly positive length. As remarked in 1.9, there can be at most finitely many jumps of magnitude exceeding $\varepsilon > 0$ during a bounded time interval. Thus, intricacies of paths are due to the intensity of jumps of small magnitude. To concentrate on those essential issues, the next construction is for processes whose jumps are all small in magnitude, say, all less than unity. We write $\mathbb{B}$ for the unit ball in $\mathbb{R}^d$ and $\mathbb{B}_\varepsilon$ for the complement in $\mathbb{B}$ of the ball of radius ε, that is,

1.22 $$\mathbb{B} = \{\, x \in \mathbb{R}^d : |x| \le 1 \,\}, \quad \mathbb{B}_\varepsilon = \{\, x \in \mathbb{R}^d : \varepsilon < |x| \le 1 \,\}.$$

1.23 THEOREM. *Let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{B}$ with mean $Leb \times \lambda$, where the measure λ on $\mathbb{B}$ satisfies $\lambda\{0\} = 0$ and*

1.24 $$\int_{\mathbb{B}} \lambda(dx)|x|^2 < \infty.$$

For ε in $(0,1)$, define

1.25 $$X_t^\varepsilon(\omega) = \int_{[0,t]\times\mathbb{B}_\varepsilon} M_\omega(ds,dx)x - t\int_{\mathbb{B}_\varepsilon} \lambda(dx)x, \quad \omega \in \Omega, \quad t \in \mathbb{R}_+.$$

Then, there exists a Lévy process X such that, for almost every ω,

$$\lim_{\varepsilon \downarrow 0} X_t^\varepsilon(\omega) = X_t(\omega),$$

the convergence being uniform in t over bounded intervals. The characteristic exponent for X is

1.26 $$\psi(r) = \int_{\mathbb{B}} \lambda(dx)(e^{ir\cdot x} - 1 - ir\cdot x), \quad r \in \mathbb{R}^d.$$

1.27 NOTATION. For future purposes, it is convenient to write

$$X_t = \int_{[0,t]\times\mathbb{B}} [M(ds, dx) - ds\lambda(dx)]\; x,$$

the exact meaning of the right side being the almost sure limit described in the preceding theorem.

1.28 REMARKS. The proof of the preceding theorem is left to the end of this section because of its length and technical nature. For the present, here are some comments on its meaning.

a) The process X^ε has the form

$$X_t^\varepsilon = Y_t^\varepsilon - a_\varepsilon t$$

where Y^ε is a compound Poisson process in $\mathbb{R}^d$ and the drift rate a_ε is a fixed vector in $\mathbb{R}^d$. To see this, we start by defining

$$b_\varepsilon = \int_{\mathbb{B}_\varepsilon} \lambda(dx)|x|, \quad c_\varepsilon = \int_{\mathbb{B}_\varepsilon} \lambda(dx)|x|^2, \qquad 0 \le \varepsilon < 1,$$

and note that $\varepsilon^2\lambda(\mathbb{B}_\varepsilon) \le \varepsilon b_\varepsilon \le c_\varepsilon \le c_0$ since $\varepsilon^2 \le \varepsilon|x| \le |x|^2$ for x in $\mathbb{B}_\varepsilon$. The condition 1.24 means that $c_0 < \infty$, which implies that $\lambda(\mathbb{B}_\varepsilon) < \infty$ and $b_\varepsilon < \infty$ for every $\varepsilon > 0$. Since $\lambda(\mathbb{B}_\varepsilon) < \infty$, the first integral on the right side of 1.25 converges absolutely, and the second defines a vector a_ε in $\mathbb{R}^d$. Hence, the claimed form for X^ε.

b) The claim of the theorem is the existence of a Lévy process X such that, for almost every ω,

$$\lim_{\varepsilon\downarrow 0} \sup_{0\le t\le u} |X_t^\varepsilon(\omega) - X_t(\omega)| = 0$$

for every u in $\mathbb{R}_+$.

c) Recall the notation introduced in Remark (a) above. If $b_0 < \infty$, then λ satisfies 1.13, and Theorem 1.12 shows that

$$Y_t = \lim_{\varepsilon\downarrow 0} Y_t^\varepsilon = \int_{[0,t]\times\mathbb{B}} M(ds, dx)x, \quad t \in \mathbb{R}_+,$$

is a pure-jump Lévy process with Lévy measure λ. In this case,

$$a = \lim_{\varepsilon\downarrow 0} \int_{\mathbb{B}_\varepsilon} \lambda(dx)\, x = \int_{\mathbb{B}} \lambda(dx)\, x$$

is also well-defined, and we have

$$X_t = Y_t - at, \quad t \in \mathbb{R}_+.$$

d) The novelty of the theorem, therefore, occurs when $b_0 = +\infty$ and $c_0 < \infty$, that is, 1.13 fails but 1.24 holds. Then, a_ε fails to converge as $\varepsilon \to 0$, and Y_t^ε fails to converge as $\varepsilon \to 0$, but the difference $X_t^\varepsilon = Y_t^\varepsilon - a_\varepsilon t$ converges. The limit process X has infinite variation over every time interval (s, t), however small $t - s > 0$ may be.

e) Every X_t^ε is a compensated sum of jumps: the sum of the sizes of jumps during $(0, t]$ is equal to Y_t^ε, the corresponding compensator term is equal to $a_\varepsilon t$, and the resulting process X^ε is a d-dimensional martingale. For this reason, the limit X is said to be a *compensated sum of jumps.*

Construction of general Lévy processes

The next theorem introduces Lévy processes of a general nature. In Section 5, Itô-Lévy decomposition theorem will show that, conversely, every Lévy process in $\mathbb{R}^d$ has this form. In the next section, there are several concrete examples.

Recall the notation $\mathbb{B}$ for the closed unit ball in $\mathbb{R}^d$, and write $\mathbb{B}^c$ for its complement, $\mathbb{R}^d \backslash \mathbb{B}$. We shall use notation 1.27 again.

1.29 THEOREM. *Let b be a vector in $\mathbb{R}^d$, and c a $d \times d'$ matrix, and λ a measure on $\mathbb{R}^d$ satisfying $\lambda\{0\} = 0$ and*

1.30 $$\int_{\mathbb{R}^d} \lambda(dx)\ (|x|^2 \wedge 1) < \infty.$$

Let W be a d'-dimensional Wiener process and, independent of it, let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean Leb $\times\, \lambda$. Then,

1.31 $$X_t = bt + cW_t + \int_{[0,t]\times\mathbb{B}} [M(ds, dx) - ds\lambda(dx)]\, x + \int_{[0,t]\times\mathbb{B}^c} M(ds, dx)\, x,$$

defines a Lévy process in $\mathbb{R}^d$, and the characteristic exponent of X is, with $v = cc^T$,

1.32 $$\psi(r) = ir\cdot b - \tfrac{1}{2} r\cdot vr + \int_{\mathbb{B}} \lambda(dx)(e^{ir\cdot x} - 1 - ir\cdot x) + \int_{\mathbb{B}^c} \lambda(dx)(e^{ir\cdot x} - 1), \quad r \in \mathbb{R}^d.$$

Proof. Let X^b, X^c, X^d, X^e denote the processes defined by the four terms on the right side of 1.31 in the order they appear; then,

1.33 $$X = X^b + X^c + X^d + X^e.$$

The first term is trivially Lévy. The second, X^c, is a continuous Lévy process, since it is the product of the matrix c with the continuous Lévy process W.

The condition 1.30 is equivalent to requiring the condition 1.24 together with $\lambda(\mathbb{B}^c) < \infty$. Thus, Theorem 1.23 shows that X^d is a Lévy process. And, as remarked in 1.20, X^e is a compound Poisson process. So, all four are Lévy.

The processes X^d and X^e are independent because the traces of the Poisson random measure M on $\mathbb{R}_+ \times \mathbb{B}$ and on $\mathbb{R}_+ \times \mathbb{B}^c$ are independent. And X^c is independent of X^d and X^e by the assumed independence of W and M. Since sums of independent Lévy processes is Lévy, X is Lévy. The formula for the characteristic exponent follows from the independence of the four terms, results in Theorem 1.12 and 1.23, and the well-known formula for $\mathbb{E}\, e^{iZ}$, where $Z = r \cdot cW_t = \sum_i \sum_j r_i c_{ij} W_t^{(j)}$ is Gaussian with mean 0 and variance $(r \cdot vr)t$. □

1.34 Remarks. a) *Lévy-Khinchine formula.* This refers to the formula 1.32. If Z is an $\mathbb{R}^d$-valued infinitely divisible variable, then $\mathbb{E}\, e^{ir\cdot Z} = e^{\psi(r)}$ for some b in $\mathbb{R}^d$, some $d \times d$ symmetric positive definite matrix v, and some measure λ on $\mathbb{R}^d$ satisfying 1.30.

b) *Characteristics for X.* This refers to the triplet (b, v, λ) which determines the probability law of X.

c) *Semimartingale connection.* The decomposition 1.31–1.33 shows that X is a semimartingale (see Definition V.5.18): The drift term X^b is continuous and has locally *bounded* variation, the Gaussian term X^c is a *continuous* martingale, X^d is a *discontinuous* martingale, and X^e is a step process whose every jump *exceeds* unity in magnitude. Thus, $X^c + X^d$ is the martingale part of X, and $X^b + X^e$ the part with locally bounded variation.

The following is immediate from Theorem 1.12 for pure-jump Lévy processes, but we state it here as a special case of the last theorem.

1.35 Corollary. *In the last theorem, suppose that λ satisfies the condition 1.13. Then, the integral*

$$a = \int_{\mathbb{B}} \lambda(dx)x$$

converges absolutely, and the process X takes the form

1.36 $$X_t = (b - a)t + cW_t + \int_{[0,t]\times\mathbb{R}^d} M(ds, dx)x, \quad t \in \mathbb{R}_+,$$

with the last term defining a pure-jump Lévy process. Accordingly, the characteristic exponent becomes

$$\psi(r) = ir \cdot (b - a) - \frac{1}{2} r \cdot vr + \int_{\mathbb{R}^d} \lambda(e^{ir\cdot x} - 1), \quad r \in \mathbb{R}^d.$$

Proof. When λ satisfies 1.13, the integral defining a converges absolutely, and λ satisfies 1.30 since $|x|^2 \le |x|$ for $x \in \mathbb{B}$. So, the conclusions of the

last theorem hold. In addition, Remark 1.28c applies and $X_t^d = Y_t - at$ in the notation there. Now, writing $Y_t + X_t^e$ as one integral, we obtain 1.36 from 1.31. □

Proof of Theorem 1.23

This will be through a series of lemmas. We start with an easy extension of Kolmogorov's inequality, Lemma III.7.1, to $\mathbb{R}^d$-valued variables. This is a discrete-time result, but we state it in continuous-time format.

1.37 LEMMA. *Let $\{ Z(t) : t \in \mathbb{R}_+ \}$ be a process with state space $\mathbb{R}^d$ and $\mathbb{E}\, Z(t) = 0$ for all t. Suppose that it has independent increments. Then, for every finite set $D \subset [0,1]$ and every $\varepsilon > 0$,*

$$\mathbb{P}\{ \sup_{t\in D} |Z(t)| > \varepsilon \} \le \frac{d}{\varepsilon^2}\, \mathbb{E}\, |Z(1)|^2.$$

Proof. Let $Z^i(t)$ denote the i-coordinate of $Z(t)$. Obviously,

$$\sup_D |Z(t)|^2 = \sup_D \sum_{i=1}^d |Z^i(t)|^2 \le \sum_{i=1}^d \sup_D |Z^i(t)|^2,$$

and the left side exceeds ε^2 only if at least one term on the right exceeds ε^2/d. Thus,

$$\begin{aligned}\mathbb{P}\{ \sup_D |Z(t)| > \varepsilon \} &\le \sum_{i=1}^d \mathbb{P}\{ \sup_D |Z^i(t)| > \frac{\varepsilon}{\sqrt{d}} \} \\ &\le \sum_{i=1}^d \frac{d}{\varepsilon^2}\, \mathbb{E}\, |Z^i(1)|^2 = \frac{d}{\varepsilon^2}\, \mathbb{E}\, |Z(1)|^2,\end{aligned}$$

where Kolmogorov's inequality justifies the second inequality. □

For processes Z with right-continuous and left-limited paths, we introduce the norm

1.38 $$\|Z\| = \sup_{0\le t\le 1} |Z(t)|.$$

The following extends Kolmogorov's inequality to continuous-time processes; we state it for Lévy processes even though the stationarity of increments is not needed.

1.39 LEMMA. *Let Z be a Lévy process in $\mathbb{R}^d$ with mean 0. For every $\varepsilon > 0$,*

$$\mathbb{P}\{ \|Z\| > \varepsilon \} \le \frac{d}{\varepsilon^2}\, \mathbb{E}\, |Z(1)|^2.$$

Proof. Let $q_0, q_1, \ldots$ be an enumeration of the rational numbers in $[0,1]$. Let $D_n = \{ q_0, \ldots, q_n \}$. By the right-continuity of Z, the supremum of $|Z(t)|$

over t in D_n increases to $\|Z\|$ as $n \to \infty$. Thus, by the monotone convergence theorem,

$$\mathbb{P}\{\ \|Z\| > \varepsilon\ \} = \lim_n \mathbb{P}\{\ \sup_{t\in D_n} |Z(t)| > \varepsilon\ \};$$

and the proof is completed via Lemma 1.37 above. □

1.40 LEMMA. *Let $Z_1, \ldots, Z_m$ be processes with state space $\mathbb{R}^d$ and paths that are right-continuous and left-limited. Suppose that $Z_1, Z_2 - Z_1, \ldots, Z_m - Z_{m-1}$ are independent. Then, for every $\varepsilon > 0$,*

1.41 $$\mathbb{P}\{\ \max_{k\le m} \|Z_k\| > 3\varepsilon\ \} \le 3 \max_{k\le m} \mathbb{P}\{\ \|Z_k\| > \varepsilon\ \}.$$

Proof. Let H be the event on the left side, and let 3δ denote the right side; we need to show that

1.42 $$\mathbb{P}(H) \le 3\delta.$$

Put $Z_0 = 0$ and let $H_k = \{\max_{j\le k-1} \|Z_j\| \le 3\varepsilon < \|Z_k\|\}$ for $k = 1, \ldots, m$; these events form a partition of H. Since $\|Z_m - Z_k\| + \|Z_m\| \ge \|Z_k\|$, we have

$$H_k \cap \{\ \|Z_m - Z_k\| \le 2\varepsilon\ \} \subset H_k \cap \{\ \|Z_m\| > \varepsilon\ \}.$$

The two events on the left side are independent for each k by the assumed independence of the increments of $k \mapsto Z_k$. The union over k of the right side yields a subset of $\{\ \|Z_m\| > \varepsilon\ \}$, and the latter's probability is at most δ. Thus,

1.43 $$\sum_{k=1}^{m} \mathbb{P}(H_k)\, \mathbb{P}\{\ \|Z_m - Z_k\| \le 2\varepsilon\ \} \le \delta.$$

Since $\|Z_m - Z_k\| \le \|Z_m\| + \|Z_k\|$, on the set $\{\ \|Z_m - Z_k\| > 2\varepsilon\ \}$ we have either $\|Z_m\| > \varepsilon$ or $\|Z_k\| > \varepsilon$. Hence,

$$1 - \mathbb{P}\{\ \|Z_m - Z_k\| \le 2\varepsilon\ \} \le \mathbb{P}\{\|Z_m\| > \varepsilon\} + \mathbb{P}\{\|Z_k\| > \varepsilon\} \le 2\delta.$$

Putting this into 1.43 and recalling that (H_k) is a partition of H, we get

$$(1 - 2\delta)\mathbb{P}(H) \le \delta.$$

If $\delta < 1/3$, then $1 - 2\delta \ge 1/3$ and we get $\mathbb{P}(H)/3 \le \delta$ as needed to show 1.42. If $\delta \ge 1/3$, then 1.42 is true trivially. □

Proof of Theorem 1.23

Recall the setup and assumptions of the theorem. Recall the norm 1.38. Let (ε_n) be a sequence in $(0, 1)$ strictly decreasing to 0. For notational simplicity, we define

1.44 $$B_n = \mathbb{B}_{\varepsilon_n}, \quad Z_n(t) = X_t^{\varepsilon_n} = \int_{[0,t]\times B_n} M(ds, dx)x - t \int_{B_n} \lambda(dx)x.$$

We shall show that, almost surely,

1.45 $$\lim_{n\to\infty} \sup_{i,j\ge n} \|Z_i - Z_j\| = 0.$$

Assuming this, the rest of the proof is as follows: 1.45 means that (Z_n) is Cauchy for almost sure convergence in the norm $\|\cdot\|$. Hence, there is a process X such that $\|Z_n - X\| \to 0$ almost surely, and it is obvious that the limit X does not depend on the sequence (ε_n) chosen. So, in the notation of Theorem 1.23, we see that, for almost every ω, $X_t^\varepsilon(\omega) \mapsto X_t(\omega)$ uniformly in $t \le 1$ as $\varepsilon \to 0$. The uniformity of convergence implies that $X(\omega)$ is right-continuous and left limited on the interval $[0,1]$, since each X^ε is such. Since almost sure convergence implies convergence in distribution for $(X_{t_1}^\varepsilon, \ldots, X_{t_k}^\varepsilon)$, and since X^ε has stationary and independent increments, the process X has stationary and independent increments, over $[0,1]$. Repeating the whole procedure for the processes $\{\ X_{k+t}^\varepsilon - X_k^\varepsilon\ :\ 0 \le t \le 1\ \}$ with $k = 1, 2, \ldots$ completes the proof of the theorem, except for showing 1.45.

Each Z_n defined in 1.44 is a Lévy process with $\mathbb{E}\ Z_n(t) = 0$. Moreover, the processes $Z_1, Z_2 - Z_1, \ldots$ are independent (and Lévy), because they are defined by the traces of M over the disjoint sets $\mathbb{R}_+ \times B_1, \mathbb{R}_+ \times (B_2 \backslash B_1), \cdots$ respectively, and M is Poisson.

Fix $\varepsilon > 0$. Applying Lemma 1.40 with processes $Z_{n+1} - Z_n, \ldots, Z_{n+m} - Z_n$ and then using Lemma 1.39 with well-known formulas for the moments of Poisson integrals, we obtain

$$\begin{aligned}
\mathbb{P}\{\ \max_{k\le m} \|Z_{n+k} - Z_n\| > 3\varepsilon\ \} &\le 3 \max_{k\le m}\ \mathbb{P}\{\ \|Z_{n+k} - Z_n\| > \varepsilon\ \} \\
&\le 3 \max_{k\le m}\ \frac{d}{\varepsilon^2}\ \mathbb{E}\ |Z_{n+k}(1) - Z_n(1)|^2 \\
&\le \frac{3d}{\varepsilon^2}\ \max_{k\le m} \int_{B_{n+k}\backslash B_n} \lambda(dx)|x|^2 \\
&\le \frac{3d}{\varepsilon^2} \int_{B_0\backslash B_n} \lambda(dx)|x|^2
\end{aligned}$$

On the left, the random variable involved increases as m does, and the limit dominates $\frac{1}{2}\|Z_i - Z_j\|$ for all $i, j \ge n$. Thus,

$$\mathbb{P}\{\ \sup_{i,j\ge n} \|Z_i - Z_j\| > 6\varepsilon\ \} \le \frac{3d}{\varepsilon^2} \int_{B_0\backslash B_n} \lambda(dx)|x|^2.$$

On the right side, the integrability condition 1.24 allows the use of the dominated convergence theorem as $n \to \infty$, and the limit is 0 since $B_0 \backslash B_n$ shrinks to the empty set. Hence, since the supremum over $i, j \ge n$ decreases as n increases,

$$\mathbb{P}\{\ \lim_n \sup_{i,j\ge n} \|Z_i - Z_j\| > 6\varepsilon\ \} = \lim_n \mathbb{P}\{\ \sup_{i,j\ge n} \|Z_i - Z_j\| > 6\varepsilon\ \} = 0.$$

Since $\varepsilon > 0$ is arbitrary, this proves that 1.45 holds almost surely.

Exercises and complements

1.46 *Simple random walk in continuous time.* Let X be a pure-jump Lévy process in $\mathbb{R}$ with Lévy measure

$$\lambda = a\delta_1 + b\delta_{-1},$$

where δ_x is Dirac at x, and a and b are positive numbers. Show that $X = X^+ - X^-$ where X^+ and X^- are independent Poisson processes with respective rates a and b. Describe the total variation process V. Show that at every time of jump for V, the process X jumps either upward or downward with respective probabilities $a/(a+b)$ and $b/(a+b)$.

1.47 *Processes with discrete jump size.* Let λ be a purely atomic finite measure in $\mathbb{R}^d$. Let X be a compound Poisson process with λ as its Lévy measure. Show that X can be decomposed as

$$X = \sum_1^\infty a_k N^{(k)}$$

where (a_k) is a sequence in $\mathbb{R}^d$, and the $N^{(k)}$ are independent Poisson processes. Identify the a_k and the rates of the $N^{(k)}$.

1.48 *Two-sided gamma processes.* As in Example 1.21, let X^+ and X^- be independent gamma processes and define $X = X^+ - X^-$. Suppose that X^+ and X^- have the same scale parameter c, and respective shape rates a and b. Let $V = X^+ + X^-$.

a) Compute the Lévy measures of X and V.
b) Show that the distribution π_t of X_t is given by

$$\pi_t f = \int_{\mathbb{R}_+} dx \frac{e^{-cx} c^{at} x^{at-1}}{\Gamma(at)} \int_{\mathbb{R}_+} dx \frac{e^{-cx} c^{bt} x^{bt-1}}{\Gamma(bt)} f(x-y), \quad f \in \mathcal{B}(\mathbb{R}).$$

c) For fixed t, show that X_t^+/V_t and V_t are independent. What are their distributions?

1.49 *Symmetric gamma distribution.* Let k_a denote the density function of the symmetric gamma distribution with shape index a and scale parameter 1, that is,

$$\int_{\mathbb{R}} dx\, k_a(x)\, e^{irx} = \left(\frac{1}{1+r^2}\right)^a, \qquad r \in \mathbb{R}.$$

a) The density for the same distribution with scale parameter c is the function $x \mapsto c k_a(cx)$. Show.
b) For $a = b$ in 1.48, show that $\pi_t(dx) = c k_{at}(cx)\, dx$.

1.50 *Alternative constructions.* Let N be a standard Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ (with mean $Leb \times Leb$). Let $j : \mathbb{R}_+ \mapsto \mathbb{R}^d$ be a Borel function satisfying

$$\int_{\mathbb{R}_+} dx\,(|j(x)| \wedge 1) < \infty.$$

a) Show that

$$X_t = \int_{[0,t]\times\mathbb{R}_+} N(ds,dx) j(x), \quad t \in \mathbb{R}_+,$$

defines a pure-jump Lévy process.

b) Compute the Lévy measure corresponding to $j(x) = e^{-cx}$, $x \in \mathbb{R}_+$.

1.51 *Continuation.* Let N be as in the preceding exercise. Let $j : \mathbb{R}_+ \mapsto \mathbb{R}^d$ be such that

$$\int_{\mathbb{R}_+} dx\,(|j(x)|^2 \wedge 1) < \infty,$$

and put $D = \{\, x \in \mathbb{R}_+ \,:\, |j(x)| \le 1 \,\}$ and $D^c = \mathbb{R}_+ \backslash D$. Let

$$X_t^d = \int_{[0,t]\times D} [\,N(ds,dx) - ds\,dx\,]\,j(x),$$

with the exact meaning to be in accord with Notation 1.27. Then, X^d is a Lévy process. So is

$$X_t^e = \int_{[0,t]\times D^c} N(ds,dx)\,j(x).$$

1.52 *Continuation.* Let j_1 and j_2 be Borel functions from $\mathbb{R}_+$ into $\mathbb{R}$, and suppose that they both satisfy the condition on j of 1.50. Define

$$X_t^{(1)} = \int_{[0,t]\times\mathbb{R}_+} N(ds,dx)\,j_1(x), \quad X_t^{(2)} = \int_{[0,t]\times\mathbb{R}_+} N(ds,dx)\,j_2(x).$$

Show that $X^{(1)}$ and $X^{(2)}$ are Lévy processes in $\mathbb{R}$, and $X = (X^{(1)}, X^{(2)})$ is a Lévy process in $\mathbb{R}^2$; all three are of the pure-jump type; $X^{(1)}$ and $X^{(2)}$ are dependent.

1.53 *Spherical coordinates.* Each point x in $\mathbb{R}^d$ can be represented as $x = vu$ by letting $v = |x|$ and $u = x/|x|$; obviously, v is the length of x, and u is its direction represented as a point on the unit sphere

$$S = \{\, x \in \mathbb{R}^d \,:\, |x| = 1 \,\}.$$

Let ρ be a σ-finite measure on $\mathbb{R}_+$ and let σ be a transition probability kernel from $\mathbb{R}_+$ into S. Define the measure λ on $\mathbb{R}^d$ by the integral formula

$$\lambda f = \int_{\mathbb{R}^d} \lambda(dx)\,f(x) \;=\; \int_{\mathbb{R}_+} \rho(dv) \int_S \sigma(v,du) f(vu)$$

for $f : \mathbb{R}^d \mapsto \mathbb{R}_+$ Borel. Then, ρ is called the radial part of λ, and σ the spherical part.

a) Show that $\int_{\mathbb{R}^d} \lambda(dx)\,(|x|^2 \wedge 1) < \infty \;\Leftrightarrow\; \int_{\mathbb{R}_+} \rho(dv)(v^2 \wedge 1) < \infty$.

b) Show that $\int_{\mathbb{R}^d} \lambda(dx)\,(|x| \wedge 1) < \infty \;\Leftrightarrow\; \int_{\mathbb{R}_+} \rho(dv)(v \wedge 1) < \infty$.

c) Let $h : \mathbb{R}^d \mapsto \mathbb{R}_+$ be the mapping $x \mapsto |x|$. Show that $\rho = \lambda \circ h^{-1}$. If λ is given somehow, one can find ρ and σ such that ρ is the radial part and σ the spherical part.

1.54 *Continuation.* Let λ, ρ, σ be as in the preceding exercise 1.53, and suppose that ρ-integral of $(v \wedge 1)$ is finite as in part (b) of 1.53. Let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $Leb \times \rho$. Let (T_i, V_i), $i \in \mathbb{N}$, be a labeling of its atoms. For each i, let U_i be a random point on the sphere S such that

$$\mathbb{P}\{\, U_i \in B \mid T_i = t, V_i = v \,\} = \sigma(v, B)$$

free of t, and assume that U_i is conditionally independent of $\{\, (T_j, V_j, U_j) \;:\; j \neq i \,\}$ given V_i. Show that

$$X_t = \sum_{i \in \mathbb{N}} V_i\, U_i\, 1_{\{T_i \le 1\}}, \quad t \in \mathbb{R}_+,$$

defines a pure-jump Lévy process X in $\mathbb{R}^d$ whose Lévy measure is λ.

2 Stable Processes

Stable processes form an important subclass of Lévy processes. This section is to introduce them and point out the explicit forms of their characteristic exponents and Lévy measures. Section 6 on subordination will have further results clarifying the relationships among them.

Let a be a number in $\mathbb{R}_+$. Let $X = (X_t)_{t \in \mathbb{R}_+}$ be a Lévy process in $\mathbb{R}^d$. Then X is said to be a-stable, or *stable with index* a, or *self-similar with index* a if the process $\hat{X} = (s^{-1/a} X_{st})_{t \in \mathbb{R}_+}$ has the same probability law as X for every s in $(0, \infty)$. Since the law of a Lévy process X is determined by the distribution of X_1, and since $\hat{X}$ is also Lévy, the condition of a-stability is equivalent to the condition that $s^{-1/a} X_s$ have the same distribution as X_1 for every s in $(0, \infty)$, or that X_t and $t^{1/a} X_1$ have the same distribution.

If $X = 0$ almost surely, then it is a-stable for every a in $\mathbb{R}_+$; we exclude this degenerate case from now on; then $a > 0$ necessarily. Exercises 2.34 and 2.35 show that the index a cannot exceed 2. If $X = W$ or $X = cW$ with W Wiener and c a constant, then X is stable with index 2; see Exercise 2.36. All other stable processes have indices in the interval $(0, 2)$.

For stable processes in $\mathbb{R}$, we shall see the following. If the index a is in $(0, 1)$, then the process is necessarily a pure-jump Lévy process whose Lévy measure is infinite and has a specific form. If a is in $(1, 2)$, then the Lévy measure is again infinite and has a specific form, and the paths have

infinite variation over every time interval and cannot be pure-jump type. If $a = 1$, there are three possibilities: the process can be pure drift and thus deterministic; or it can be a Cauchy process, the paths having the same qualitative features as in the case of indices in $(1, 2)$, but each increment having a Cauchy distribution; or it can be a Cauchy process plus some drift.

Stable processes with index in $(0, 1)$

The process introduced in Example VI.4.10 is an increasing pure-jump Lévy process which is stable with index a in $(0, 1)$. It will serve as the total variation process (see 1.12 *et seq.*) for a-stable processes in $\mathbb{R}^d$. We review the example in a form suited to our current agenda.

2.1 EXAMPLE. *Increasing stable processes.* Fix a in $(0, 1)$ and c in $(0, \infty)$. Let

$$\lambda(dv) = dv\ \frac{c}{v^{a+1}}\ 1_{(0,\infty)}(v), \quad v \in \mathbb{R}.$$

This λ satisfies the condition 1.13 of Theorem 1.12. Let V be the pure-jump Lévy process associated. Then V is strictly increasing, all its jumps are upward, and it has infinitely many jumps in every time interval of some length; the last is because $\lambda(0, \infty) = +\infty$. The process V is a-stable; conversely, every increasing stable process has this form. Recall from VI.4.10 that, for $p \geq 0$,

2.2 $$\mathbb{E}\, e^{\,-pV_t} = \exp_-\ t \int_{\mathbb{R}_+} dv\ \frac{c}{v^{a+1}}\ (1 - e^{-pv}) = \exp_-\ tc\ \frac{\Gamma(1-a)}{a}\ p^a.$$

We show next that the corresponding characteristic function is

$$\mathbb{E}\, e^{irV_t} = \exp t \int_{\mathbb{R}_+} dv\ \frac{c}{v^{a+1}} (e^{irv} - 1)$$

2.3 $$= \exp_-\ tc_a |r|^a\ [1 - i(\tan \tfrac{1}{2}\pi a)\ \operatorname{sgn} r],$$

$r \in \mathbb{R}$, where

2.4 $$c_a\ =\ c\ \frac{\Gamma(1-a)}{a} \cos \tfrac{1}{2}\pi a, \quad \operatorname{sgn} r\ =\ 1_{\mathbb{R}_+}(r) - 1_{\mathbb{R}_+}(-r).$$

We start by claiming that, for every complex number z whose real part is zero or less,

$$\int_{\mathbb{R}_+} dv \frac{c}{v^{a+1}} (e^{zv} - 1)\ =\ -c\frac{\Gamma(1-a)}{a}\ (-z)^a\ =\ -c\frac{\Gamma(1-a)}{a} |z|^a\ e^{ia\ \operatorname{Arg}(-z)},$$

2.5

where $\operatorname{Arg} z$ is the principal value (in the interval $(-\pi, \pi]$) of the argument of z. This claim follows from noting that 2.5 holds for all negative real z in view of 2.2, and that both sides of 2.5 are regular in the interior of the left-hand plane and are continuous on the boundary. Taking $z = ir$ in 2.5 we obtain 2.3, since $\operatorname{Arg}(-ir) = -\frac{1}{2}\pi\ \operatorname{sgn} r$.

2.6 Example. *Stable processes in $\mathbb{R}^d$ with index in* $(0,1)$. Fix a in $(0,1)$ and c in $(0,\infty)$. Let $S = \{ x \in \mathbb{R}^d : |x| = 1 \}$, the unit sphere in $\mathbb{R}^d$, and let σ be a probability measure on it. Define a measure λ on $\mathbb{R}^d$ by the integral formula

$$2.7 \quad \lambda f = \int_{\mathbb{R}^d} \lambda(dx)\, f(x) = \int_{\mathbb{R}_+} dv\, \frac{c}{v^{a+1}} \int_S \sigma(du)\, f(vu), \qquad f \geq 0 \text{ Borel};$$

see Exercise 1.53; note that the radial part of λ is the Lévy measure of the preceding example. When $f(x) = |x| \wedge 1$, we have $f(vu) = |vu| \wedge 1 = v \wedge 1$ for u in S, and it follows that $\lambda f < \infty$ since $a \in (0,1)$. Thus λ satisfies the condition 1.13.

Let X be the process constructed in Theorem 1.12 for this Lévy measure λ. Then, X is a pure-jump Lévy process in $\mathbb{R}^d$. Its total variation process V is the increasing pure-jump Lévy process of the preceding example. The processes X and V have the same jump times, infinitely many in every open interval. Moreover, 2.7 implies the following conditional structure for X given V: given that V has a jump of size v at time t, the process X has a jump of size vU at the same time t, where U is a random variable with distribution σ on the sphere S ; see Exercise 1.54 for the same description in more detail.

Consequently, the a-stability of V implies the a-stability of X: for fixed s in $(0,\infty)$, the transformation that takes the sample path $t \mapsto X_t(\omega)$ to the sample path $t \mapsto \hat{X}_t(\omega) = s^{-1/a} X_{st}(\omega)$ merely alters the times and magnitudes of the jumps, which are totally determined by $t \mapsto V_t(\omega)$. The a-stability of X can also be deduced by noting that $t^{1/a}X_1$ and X_t have the same characteristic function; for, with the notations 2.4, it follows from 2.3 and 2.7 that, with c_a as in 2.4,

$$2.8 \quad \mathbb{E}\, e^{ir\cdot X_t} = \exp t \int_S \sigma(du) \int_{\mathbb{R}_+} dv \frac{c}{v^{a+1}} (e^{ivr\cdot u} - 1)$$

$$= \exp_- \; tc_a \int_S \sigma(du) |r\cdot u|^a [1 - i(\tan \tfrac{1}{2}\pi a)\, \text{sgn} \quad r\cdot u],\ r \in \mathbb{R}^d.$$

2.9 Example. *Symmetric stable processes with index in* $(0,1)$. A Lévy process X is said to be *symmetric* if $-X$ has the same law as X. This is equivalent to having the characteristic exponent ψ symmetric, that is, to having $\psi(r) = \psi(-r)$ for every r in $\mathbb{R}^d$. In the case of a pure-jump Lévy process, in view of 1.14 defining X, symmetry is equivalent to having the law of M invariant under the transformation $(t,x) \mapsto (t,-x)$ of $\mathbb{R}_+ \times \mathbb{R}^d$ onto itself. These imply, together with 2.7 and 2.8, that the following four statements are equivalent for the process X of Example 2.6:

a) The process X is symmetric.

b) The Lévy measure λ is symmetric, that is, $\lambda(B) = \lambda(-B)$ for Borel $B \subset \mathbb{R}^d$.

c) The distribution σ is symmetric, that is, $\sigma(B) = \sigma(-B)$ for Borel $B \subset S$.

d) The exponent of X is real-valued, that is, 2.8 reduces to

2.10 $$\mathbb{E}\, e^{\,ir\cdot X_t} \;=\; \exp_-\; tc_a \int_S \sigma(du)\; |r\cdot u|^a, \quad r\in\mathbb{R}^d.$$

2.11 EXAMPLE. *Isotropic stable processes with index in* $(0,1)$. A Lévy process X in $\mathbb{R}^d$ is said to be *isotropic*, or *rotationally invariant*, if its law is invariant under all orthogonal transformations of $\mathbb{R}^d$. This is equivalent to saying that X and gX have the same law for every orthogonal matrix g of dimension d. If $d=1$, isotropy is the same as symmetry; in higher dimensions, isotropy implies symmetry and more.

Let X be as in Example 2.6. Thinking of the jumps, it is clear that X is isotropic if and only if the law governing the jump directions is isotropic, that is, the measure σ on S is the uniform distribution on S. And then, the characteristic function 2.8 becomes even more specific than 2.10 (see Exercises 2.40 and 2.41 for the computations):

2.12 $$\mathbb{E}\, e^{\,ir\cdot X_t} \;=\; \exp_-\; tc_{ad}\; |r|^a, \quad r\in\mathbb{R}^d,$$

where the constant c_{ad} depends on a, c, and d; with c_a as in 2.4,

2.13 $$c_{ad} \;=\; \frac{\Gamma(\frac{a+1}{2})\Gamma(\frac{d}{2})}{\Gamma(\frac{a+d}{2})\Gamma(\frac{1}{2})}\, c_a.$$

Stable processes with index 1

If a Lévy process X is 1-stable, then X_t and tX_1 have the same distribution. The meaning of stability is striking: X_5, for instance, which is the sum of 5 independent copies of X_1, has the same distribution as $5X_1$. The simplest example is the pure-drift process $X_t = bt, t\in\mathbb{R}_+$. If X is 1-stable, then so is $\hat{X} = (X_t + bt)$; but, if $\hat{X}$ is to be symmetric, X has to be symmetric and $b=0$. From now on we concentrate on processes without drift.

2.14 EXAMPLE. *Standard Cauchy process in* $\mathbb{R}$. This is a symmetric stable process with index 1. Its canonical decomposition 1.33 is $X = X^d + X^e$ and the Lévy measure defining its law is

2.15 $$\lambda(dx) \;=\; dx\, \frac{1}{\pi x^2}, \quad x\in\mathbb{R}.$$

This Lévy measure satisfies 1.30 but not 1.13. We shall show that

2.16 $$\mathbb{E}\, e^{\,irX_t} \;=\; e^{-t\,|r|}, \quad r\in\mathbb{R},$$

which makes it apparent that X is 1-stable. The corresponding distribution is

2.17 $$\mathbb{P}\{\, X_t \in dx \,\} \;=\; dx\, \frac{t}{\pi(t^2+x^2)}, \quad x\in\mathbb{R},$$

which is called the *Cauchy distribution* with scale parameter t, because it is the distribution of tX_1 and the distribution of X_1 is the standard Cauchy distribution; see II.2.27. For this reason, X is said to be a standard Cauchy process.

The symmetry of λ simplifies the construction of X. Going over Theorems 1.23 and 1.29, we observe that the λ-integral in 1.25 vanishes and thus the term X^d is a limit of pure-jump (compound Poisson) processes. Thus, $X = X^d + X^e$ can be written as

2.18 $$X_t = \int_{[0,t]\times\mathbb{R}} M(ds, dx)\ x,$$

the precise meaning of which is as follows: with $\mathbb{R}_\varepsilon = \mathbb{R} \setminus (-\varepsilon, \varepsilon)$, for almost every ω,

2.19 $$X_t(\omega) = \lim_{\varepsilon \downarrow 0} \int_{[0,t]\times\mathbb{R}_\varepsilon} M_\omega(ds, dx)x,$$

the convergence being uniform in t over bounded intervals. It follows from this, or from 1.32 and the mentioned vanishing of the λ-integral on the right side of 1.25, that

$$\begin{aligned}\mathbb{E}\ e^{\ irX_t} &= \exp\ t \lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}_\varepsilon} \lambda(dx)\ (e^{irx} - 1)\\ &= \exp_-\ 2t \int_{\mathbb{R}_+} dx\ \frac{1}{\pi x^2}\ (1 - \cos rx) = e^{-t|r|},\ r \in \mathbb{R},\end{aligned}$$

as claimed in 2.16.

The Cauchy process X is not a pure-jump process despite the looks of 2.18. Indeed, since

$$\int_{(0,1)} \lambda(dx)\ x = \int_{(-1,0)} \lambda(dx)\ (-x) = +\infty,$$

it follows from Proposition VI.2.13 on the finiteness of Poisson integrals that

$$\int_{(s,t)\times(0,1)} M_\omega(du, dx)\ x = \int_{(s,t)\times(-1,0)} M_\omega(du, dx)\ (-x) = +\infty$$

for almost every ω for $s < t$. In other words, over every interval (s, t), the path $X(\omega)$ has infinitely many upward jumps whose sizes sum to $+\infty$, and infinitely many downward jumps whose sizes sum to $-\infty$. In particular, the total variation over (s, t) is equal to $+\infty$ always.

Nevertheless, the small jumps are small enough, and the positive and negative jumps balance each other well, that removing the big jumps yields a martingale. Employing a notation similar to 2.18 with precise meaning analogous to 2.19,

$$Z_t = \int_{[0,t]\times[-b,b]} M(ds, dx)\ x, \quad t \in \mathbb{R}_+,$$

defines a martingale Z for each b in $(0, \infty)$. Indeed, Z is a Lévy process with $\mathbb{E} Z_t = 0$ and $\mathrm{Var} Z_t = 2bt/\pi$; it is not a stable process.

The process X is not a martingale for the simple reason that $\mathbb{E} X_t$ does not exist, which is because the jumps exceeding b in magnitude are very big in expectation:

$$\mathbb{E} \int_{[0,t]\times(b,\infty)} M(ds, dx)\, x \;=\; t \int_{(b,\infty)} \lambda(dx)\, x \;=\; \infty\,,$$

and similarly for the integral over $[0,t] \times (-\infty, -b)$. This fact about $\mathbb{E} X_t$ is often expressed by saying that the Cauchy distribution has fat tails; the account above is more revealing.

2.20 EXAMPLE. *Half-Cauchy.* This is a Lévy process X that is not stable, and the distribution of X_t is not Cauchy. We give it here to clarify the role of symmetry in the 1-stability of Cauchy processes. Let X be a Lévy process in $\mathbb{R}$ whose canonical decomposition is $X = X^d + X^e$ and whose Lévy measure is

$$\lambda(dx) \;=\; dx\, \frac{1}{x^2}\, 1_{(0,\infty)}(x), \quad x \in \mathbb{R}.$$

This λ is, up to a constant multiple, the one-sided version of the Lévy measure in the preceding example.

All jumps of X are upward, but X is not constrained to $\mathbb{R}_+$; for $t > 0$, the distribution of X_t puts strictly positive mass on every interval (x, y) with $-\infty < x < y < \infty$. In particular, all the jumps of X^d are upward, the jumps over (s, t) are infinitely many and their sizes sum to $+\infty$. Thus, X^d is truly a compensated sum of jumps; it is the limit of the processes X^ε with upward jumps and downward drift.

The characteristic function for X_t is, in view of 1.32, and the form of λ here,

2.21
$$\begin{aligned}\mathbb{E}\, e^{irX_t} &= \exp t \left[\int_0^1 dx \frac{1}{x^2} (e^{irx} - 1 - irx) + \int_1^\infty dx \frac{1}{x^2} (e^{irx} - 1) \right] \\ &= \exp_- \; t \; \left[\tfrac{1}{2}\pi |r| - ic_0 r + ir \log |r| \right],\end{aligned}$$

where

$$c_0 \;=\; \int_0^1 dx\, \frac{1}{x^2}\, (\sin x - x) + \int_1^\infty dx\, \frac{1}{x^2}\, \sin x.$$

It is checked easily that it is impossible for X_t and $t^{1/a} X_1$ to have the same characteristic function for some $a > 0$. So, X is not stable at all.

2.22 EXAMPLE. *Cauchy and other* 1*-stable processes in* $\mathbb{R}^d$. Let S be the unit sphere in $\mathbb{R}^d$, and let σ be a probability measure on S satisfying

2.23
$$\int_S \sigma(du)\, u \;=\; 0.$$

For example, if $d = 2$, we obtain such a measure by putting equal weights at the vertices of a regular pentagon circumscribed by the unit circle S.

Let c be a constant in $(0, \infty)$, and let λ be the measure on $\mathbb{R}^d$ given by

2.24 $$\lambda f = \int_{\mathbb{R}^d} \lambda(dx)\, f(x) = \int_{\mathbb{R}_+} dv\, \frac{c}{v^2} \int_S \sigma(du)\, f(vu), \quad f \geq 0 \text{ Borel.}$$

This λ satisfies 1.30; for $f(x) = |x|^2 \wedge 1$, we get $\lambda f = 2c < \infty$. But λ fails to satisfy 1.13.

Let X be the Lévy process whose canonical decomposition is $X = X^d + X^e$ in the notations of 1.31 and 1.33 and whose Lévy measure is the current λ. Its sample path behavior is similar to that of Example 2.6, except that it is not a pure-jump process and has infinite variation over every interval. The magnitudes of its jumps are regulated by the radial part of λ, and the latter is a constant multiple of the Lévy measure in Example 2.20, the half-Cauchy process. The characteristic function of X_t can be obtained using 2.21:

2.25 $$\begin{aligned} \mathbb{E}\, e^{\,ir\cdot X_t} &= \exp\ t \int_S \sigma(du) \int_{\mathbb{R}_+} dv\, \frac{c}{v^2}\, [\, e^{ivr\cdot u} - 1 - ivr\cdot u\ 1_{\mathbb{B}}(vu)\,] \\ &= \exp_\ \ tc \int_S \sigma(du)\ [\, \tfrac{1}{2}\pi\ |r\cdot u| - ic_0 r\cdot u + ir\cdot u \log |r\cdot u|\,] \\ &= \exp_\ \ tc \int_S \sigma(du)\ [\, \tfrac{1}{2}\pi |r\cdot u| + ir\cdot u \log |r\cdot u|\,]\ ; \end{aligned}$$

here, we noted that $1_{\mathbb{B}}(vu) = 1_{[0,1]}(v)$ for the unit ball $\mathbb{B}$ and the unit vector u, and then used 2.21 with r there replaced by $r \cdot u$ and finally the assumption 2.23. Replacing r by tr and using 2.23 once more, we see that the characteristic functions of X_t and tX_1 are the same. Hence, X is 1-stable.

When $d = 1$, the unit "sphere" consists of the two points $+1$ and -1, and the condition 2.23 makes σ symmetric. Thus, when $d = 1$, the process is symmetric necessarily and is a Cauchy process (equal to $(1/2)\pi cZ$ where Z is standard Cauchy). In higher dimensions, symmetry and isotropy require further conditions on σ. For example, the pentagonal σ mentioned above yields a 1-stable process X in $\mathbb{R}^2$ that is not symmetric.

When $d \geq 2$, the process X is symmetric if and only if σ is symmetric, and then 2.23 holds automatically and 2.25 becomes

2.26 $$\mathbb{E}\, e^{ir\cdot X_t} = \exp_\ \tfrac{1}{2}\pi ct \int_S \sigma(du)\ |r\cdot u|, \quad r \in \mathbb{R}^d.$$

Moreover, X is isotropic if and only if σ is the uniform distribution on S, in which case the integral over the sphere can be computed as in Exercise 2.41, and we get

2.27 $$\mathbb{E}\, e^{ir\cdot X_t} = \exp_\ \hat{c}t|r|, \quad r \in \mathbb{R}^d,$$

where $\hat{c} = \frac{1}{2}c\sqrt{\pi}\,\Gamma(\frac{d}{2})/\Gamma(\frac{d+1}{2})$. This Fourier transform is invertible (see Example 6.8 for a direct computation)

$$\mathbb{P}\{X_t \in dx\} = dx\ \hat{c}t\,\Gamma\left(\tfrac{d+1}{2}\right)/[\,\pi\hat{c}^2t^2 + \pi|x|^2\,]^{(d+1)/2}, \quad x \in \mathbb{R}^d.$$

This is called the d-dimensional *Cauchy distribution* with scale factor $\hat{c}t$; thus, X_t has the same distribution as $\hat{c}tZ$, where Z has the standard d-dimensional Cauchy distribution; see Exercise 2.42 for the definition.

Stable processes with index in $(1, 2)$

These processes are similar to the stable ones with index in $(0, 1)$, except that they cannot have bounded variation over intervals.

Fix a in $(1, 2)$ and c in $(0, \infty)$. Let S be the unit sphere in $\mathbb{R}^d$, and let σ be a probability measure on it. Define a measure λ on $\mathbb{R}^d$ by

$$\lambda f \;=\; \int_{\mathbb{R}_+} dv\, \frac{c}{v^{a+1}} \int_S \sigma(du)\, f(vu), \qquad f \geq 0 \text{ Borel.}$$

This λ is the same as that in 2.7 but the shape index a is now in the interval $(1, 2)$; this λ satisfies 1.30 but not 1.13. Thus, the process we are about to introduce will have infinitely many jumps over every interval and, further, it will have infinite variation over every interval.

Theorem 1.29 shows the existence of a Lévy process $X^d + X^e$ whose Lévy measure is λ. Consider the compound Poisson process X^e whose every jump exceeds unity in magnitude; it has a well-defined mean: since $a > 1$,

$$\mathbb{E}\, X_t^e \;=\; t \int_1^\infty dv\, \frac{c}{v^{a+1}}\, v \int_S \sigma(du)\, u \;=\; t\, \frac{c}{a-1} \int_S \sigma(du)\, u \;=\; bt$$

with an apparent definition for the vector b in $\mathbb{R}^d$. We define the process X to have the canonical decomposition $X = X^b + X^d + X^e$ with $X_t^b \;=\; -bt$. In other words, in the spirit of Notation 1.27, and with M Poisson with mean $\text{Leb} \times \lambda$,

2.28 $$X_t \;=\; \int_{[0,t]\times\mathbb{R}^d} [\, M(ds, dx) - ds\, \lambda(dx)\,]\; x.$$

It is clear that X is a Lévy process in $\mathbb{R}^d$, and its every component is a martingale. It follows from 2.28 that

2.29 $$\mathbb{E}\, e^{\,ir\cdot X_t} = \exp\, t \int_{\mathbb{R}^d} \lambda(dx)\, (e^{ir\cdot x} - 1 - ir\cdot x)$$
$$= \exp\, t \int_S \sigma(du) \int_{\mathbb{R}_+} dv\, \frac{c}{v^{a+1}}\, (e^{ivr\cdot u} - 1 - ivr\cdot u).$$

It is now easy to check that X_t and $t^{1/a} X_1$ have the same characteristic function; thus, X is a-stable.

On the right side of 2.29, the integral over $\mathbb{R}_+$ can be evaluated through integration by parts using 2.3. The result is similar to 2.8: for r in $\mathbb{R}^d$,

2.30 $$\mathbb{E}\, e^{\,ir\cdot X_t} \;=\; \exp_-\; tc_a \int_S \sigma(du)\, |r\cdot u|^a\, [1 - i\, (\tan \tfrac{1}{2}\pi a)\, \operatorname{sgn} r\cdot u],$$

where

$$c_a \;=\; -\,c\,\frac{\Gamma(2-a)}{a\;(a-1)}\;\cos\tfrac{1}{2}\pi a\;;$$

note that $c_a > 0$. The formula 2.30 shows that X is symmetric if and only if σ is symmetric, in which case

2.31 $$\mathbb{E}\, e^{\,ir\cdot X_t} \;=\; \exp_-\; tc_a \int_S \sigma(du)\;|r\cdot u|^a, \quad r\in\mathbb{R}^d.$$

Further, X is isotropic if and only if σ is the uniform distribution on S, in which case the result of Exercise 2.41 yields

2.32 $$\mathbb{E}\, e^{\,ir\cdot X_t} \;=\; \exp_-\; tc_{ad}\;|r|^a, \quad r\in\mathbb{R}^d,$$

with $c_{ad} \;=\; c_a\;\Gamma(\frac{a+1}{2})\Gamma(\frac{d}{2})\;/\;\Gamma(\frac{a+d}{2})\Gamma(\frac{1}{2})$ with c_a as in 2.30. Note that 2.32 has the same form as in 2.12 and 2.26.

Exercises

2.33 *Arithmetics*. Fix $a > 0$. Show the following for Lévy processes X and Y in $\mathbb{R}^d$.

a) If X is a-stable, then so is cX for every constant c in $\mathbb{R}$.

b) If X and Y are a-stable and independent, then $X+Y$ and $X-Y$ are a-stable.

c) If X is a-stable, and Y is independent of X and is b-stable for some $b > 0$ distinct from a, then $X+Y$ is not stable.

2.34 *Stability index*. Fix $a > 0$. Suppose that X is an a-stable non-degenerate Lévy process in $\mathbb{R}$ with characteristic exponent ψ. This is to show that, then, $a \in (0,2]$ necessarily.

a) Show that $t\psi(r) \;=\; \psi(t^{1/a}\;r)$ for $t > 0$ and r in $\mathbb{R}$. Show that $\psi(r) \;=\; c\;r^a$ for some complex constant c for r in $\mathbb{R}_+$.

b) Suppose that X is symmetric, that is, X and $-X$ have the same law. Then, $\psi(r) \;=\; \psi(-r)$ for all r. Show that $\psi(r) \;=\; c\;|r|^a$ for all r in $\mathbb{R}$.

c) Show that $e^{c|r|^a}$ cannot be a characteristic function when $a > 2$. Hint: See Exercise II.2.33 about the second moment of X_1. Conclude that, if X is symmetric, then $a \in (0,2]$.

d) If X is not symmetric, let Y be an independent copy of it. Then, $X-Y$ is symmetric and a-stable. So, $a \in (0,2]$ again.

2.35 *Continuation.* Let X be a Lévy process in $\mathbb{R}^d$. Suppose that it is not degenerate. If it is a-stable, then $a \in (0,2]$. Show.

2.36 *Stability with index* 2. Let X be a Lévy process in $\mathbb{R}$. Suppose that it is 2-stable. Then X_t has the Gaussian distribution with mean 0 and variance vt for some constant v. Show.

2.37 *Stable Poissons.* Let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with mean $\mu = \text{Leb} \times \lambda$. Let $a > 0$ be fixed. Suppose that M and $M \circ h^{-1}$ have the same law (which means that $\mu = \mu \circ h^{-1}$) for $h : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+ \times \mathbb{R}$ defined by

$$h(t, x) = (\frac{t}{s}, s^{1/a} x),$$

and that this is true for every $s > 0$. Show that, then,

$$\lambda(dx) = dx\ |x|^{-a-1}\ [\ b\ 1_{(0,\infty)}(x) + c\ 1_{(-\infty,0)}(x)\ ], \quad x \neq 0,$$

for some constants b and c in $\mathbb{R}_+$. If λ satisfies 1.13 then $a \in (0, 1)$; show. If λ satisfies 1.30, then $a \in (0, 2)$; show.

2.38 *Stable processes with index in* $(0, 1)$. Let X be as in Example 2.6, but the dimension is $d = 1$. Then, the "sphere" S consists of the two points $+1$ and -1.

a) Show that λ of 2.7 takes the form

$$\lambda(dx) = dx\ \frac{c}{|x|^{a+1}}\ (p\ 1_{(0,\infty)}(x) + q\ 1_{(-\infty,0)}(x)),$$

where p and q are positive numbers with $p + q = 1$.

b) Conclude that $X = X^+ - X^-$, where X^+ and X^- are independent a-stable increasing Lévy processes.

c) Show that the characteristic exponent of X is (see 2.8)

$$\psi(r) = -\ c_a\ |r|^a\ [\ 1 - i\ (p - q)(\tan\ \tfrac{1}{2}\pi a)\ \text{sgn}\ r\ ], \quad r \in \mathbb{R}.$$

2.39 *Continuation: symmetric case.* When $d = 1$, symmetry and isotropy are the same concept. Show that X of the preceding exercise is symmetric if and only if $p = q$. Then, the characteristic exponent becomes $\psi(r) = -c_a\ |r|^a$. Check that 2.10 and 2.12 coincide when $d = 1$.

2.40 *Uniform distribution on* S. Let σ be the uniform distribution on the unit sphere S in $\mathbb{R}^d$. This is to show that, for every s in S and a in $\mathbb{R}_+$,

$$\int_S \sigma(du)\ |s \cdot u|^a = \frac{\Gamma(\frac{a+1}{2})\Gamma(\frac{d}{2})}{\Gamma(\frac{a+d}{2})\Gamma(\frac{1}{2})}.$$

The left side is $\mathbb{E}\ |s \cdot U|^a$ where U has the uniform distribution on S. The trick is to recall that, if the random vector $Z = (Z_1, \ldots, Z_d)$ has independent components each of which has the standard Gaussian distribution, then

$$Z = RU,$$

where $R = |Z|$, and U is independent of R and has the uniform distribution on S. It follows that, for every s in S,

$$\mathbb{E}\ |s \cdot Z|^a = (\mathbb{E}\ R^a)(\mathbb{E}\ |s \cdot U|^a),$$

and the problem reduces to evaluating the expectations concerning R and $s \cdot Z$.

a) Recall that R^2 has the gamma distribution with shape index $d/2$ and scale index $1/2$. Use this to show that

$$\mathbb{E}\ R^a = \int_0^\infty dx\ \frac{e^{-x}\ x^{d/2-1}}{\Gamma(d/2)}\ (2x)^{a/2}\ =\ 2^{a/2}\ \Gamma(\frac{d+a}{2})/\Gamma(d/2).$$

b) Show that $|s \cdot Z|$ has the same distribution as R but with d put equal to 1. Thus, $\mathbb{E}\ |s \cdot Z|^a\ =\ 2^{a/2}\ \Gamma(\frac{a+1}{2})/\Gamma(1/2)$.

c) Show that $\mathbb{E}\ |s \cdot U|^a$ is as claimed.

2.41 *Continuation.* For r in $\mathbb{R}^d$ and u in S, we have $|r \cdot u| = |r|\ |s \cdot u|$ with $s = r/|r|$ in S. Use this observation to show that:

$$\int_S \sigma(du)\ |r \cdot u|^a\ =\ |r|^a\ \frac{\Gamma(\frac{a+1}{2})\Gamma(\frac{d}{2})}{\Gamma(\frac{a+d}{2})\Gamma(\frac{1}{2})}.$$

2.42 *Cauchy distribution on* $\mathbb{R}^d$. Let Z take values in $\mathbb{R}^d$. It is said to have the *standard Cauchy distribution* if

$$\mathbb{P}\{Z \in dx\}\ =\ dx\ \frac{d+1}{2}\ (\frac{1}{\pi(1+|x|^2)})^{(d+1)/2}, \quad x \in \mathbb{R}^d.$$

Then, $\mathbb{E}\ e^{ir\cdot Z} = e^{-|r|}, r \in \mathbb{R}^d$. Show that Z has the same distribution as X/Y, where $X = (X_1, \ldots, X_d)$ is a d-dimensional standard Gaussian, and Y is a one-dimensional Gaussian independent of X. Note that each component Z_i has the standard one-dimensional Cauchy distribution, but the components are dependent. Show that, for every vector v in $\mathbb{R}^d$, the inner product $v \cdot Z$ has the one-dimensional Cauchy distribution with scale factor $|v|$, that is,

$$\mathbb{P}\{\ v \cdot Z \in dx\ \}\ =\ dx\ \frac{|v|}{\pi(|v|^2 + |x|^2)}, \quad x \in \mathbb{R}.$$

2.43 *Stable processes in* $\mathbb{R}$ *with index in* $(1,2)$. Let X be defined by 2.28, but with $d = 1$. Show that the Lévy measure λ defining its law has the form

$$\lambda(dx)\ =\ dx\ \frac{c}{|x|^{a+1}}\ [\ p\ 1_{(0,\infty)}(x) + q\ 1_{(-\infty,0)}(x)\ ], \quad x \in \mathbb{R},$$

where $a \in (1,2)$ and $c \in (0,\infty)$ as before, and p and q are positive numbers with $p+q = 1$. All the jumps are upward if $p = 1$, and all downward if $q = 1$. The process is symmetric if and only if $p = q = 1/2$. In all cases, X is a martingale. In particular, $\mathbb{E}\ X_t = 0$. Compute Var X_t. Show that

$$\mathbb{E}^{e\ irX_t}\ =\ \exp_\ tc_a\ [\ |r|^a - i\ (p-q)(\tan\frac{1}{2}\ \pi a)\mathrm{sgn}\ r\ ], \quad r \in \mathbb{R}.$$

2.44 *Continuation.* Recall from 2.28 that X has the form $X = X^b + X^d + X^e$, where $X_t^b = -\mathbb{E}X_t^e = -(p-q)\frac{c}{a-1}t$. Note that none of the processes $X^b, X^d, X^e, X^d + X^e, X^e + X^b, X^b + X^d$ is a-stable.

2.45 *Continuation.* Show that it is possible to decompose X as

$$X = Y - Z,$$

where Y and Z are independent a-stable processes, with Y having only upward jumps, and Z only downward jumps. Define Y and Z carefully from the same M that defines X.

3 Lévy Processes on Standard Settings

This section is to re-introduce Lévy processes in a modern setting, show the Markov and strong Markov properties for them, and reconcile the differences from the earlier definition. The motivation for modernity is two-fold: First, we prefer filtrations that are augmented and right-continuous, because of the advantages mentioned in the last section of Chapter V. Second, it is desirable to have a moving coordinate system for time and space, which would indicate what time is meant by "present" in a given argument.

Lévy processes over a stochastic base

3.1 Definition. *A* stochastic base *is a collection*

$$\mathscr{B} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$$

where $(\Omega, \mathcal{H}, \mathbb{P})$ *is a complete probability space,* $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ *is an augmented right-continuous filtration on it, and* $\theta = (\theta_t)_{t\in\mathbb{R}_+}$ *is a semigroup of operators* $\theta_t : \omega \mapsto \theta_t\omega$ *from* Ω *into* Ω *with*

3.2 $$\theta_0\omega = \omega, \quad \theta_u(\theta_t\omega) = \theta_{t+u}\omega, \quad t, u \in \mathbb{R}_+.$$

Operators θ_t *are called* time-shifts.

3.3 Definition. *Let* $X = (X_t)_{t\in\mathbb{R}_+}$ *be a stochastic process with state space* $\mathbb{R}^d$*. It is called a* Lévy process *over the stochastic base* $\mathscr{B}$ *if* X *is adapted to* $\mathcal{F}$ *and the following hold:*

a) Regularity. X *is right-continuous and left-limited, and* $X_0 = 0$.
b) Additivity. $X_{t+u} = X_t + X_u \circ \theta_t$ *for every* t *and* u *in* $\mathbb{R}_+$.
c) Lévy property. *For every* t *and* u *in* $\mathbb{R}_+$, *the increment* $X_u \circ \theta_t$ *is independent of* $\mathcal{F}_t$ *and has the same distribution as* X_u.

Remark. If X is a Lévy process over the base $\mathscr{B}$, then it is a Lévy process in the sense of Definition 1.1 with respect to the filtration $\mathcal{F}$. The difference between Definitions 1.1 and 3.3 is slight: we shall show in Theorem 3.20 below (see also Remark 3.21) that starting from a raw Lévy process (in the sense of Definition 1.1 and with respect to its own filtration $\mathcal{G}$), we can modify $\mathcal{H}, \mathcal{G}, \mathbb{P}$ to make them fit a stochastic base. The existence of shift operators is easy to add as well; see Exercise 3.24.

Shifts

Existence of shift operators is a condition of richness on Ω. In canonical constructions, it is usual to take Ω to be the collection of all right-continuous left-limited functions $\omega : \mathbb{R}_+ \to \mathbb{R}^d$ with $\omega(0) = 0$. Then, we may define $\theta_t\omega$ to be the function $u \mapsto \omega(t+u) - \omega(t)$, and setting $X_t(\omega) = \omega(t)$ we obtain both the semigroup property 3.2 and the additivity 3.3b.

In general, θ_t derives its meaning from what it does, which is described by the additivity condition 3.3b. We interpret it as follows: $X_u \circ \theta_t$ is the increment over the next period of length u if the present time is t. Thus, θ_t shifts the time-space origin to the point (t, X_t) of the standard coordinate system; see Figure 9 below.

In other words, θ is a moving reference frame pinned on the path X. It is an egocentric coordinate system: the present is the origin of time, the present position is the origin of space.

Our usage of shifts is in accord with the established usage in the theory of Markov processes. We illustrate this with an example and draw attention to a minor distinction in terms. Let X be a Wiener process, and put

$$Z_t \;=\; Z_0 + X_t, \quad t \in \mathbb{R}_+.$$

Then, Z is called a standard Brownian motion with initial position Z_0. The established usage would require that the Markov process Z satisfy

$$Z_u \circ \theta_t \;=\; Z_{t+u}, \quad t, u \in \mathbb{R}_+ \;;$$

This is called time-homogeneity for Z. It implies that X is additive:

$$X_u \circ \theta_t \;=\; Z_u \circ \theta_t - Z_0 \circ \theta_t \;=\; Z_{t+u} - Z_t \;=\; X_{t+u} - X_t.$$

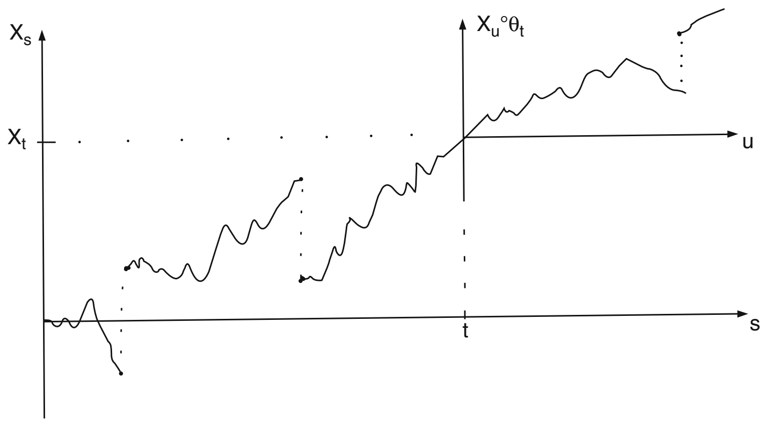


Figure 9: When the present time is t, the new coordinate system has its origin at (t, X_t) of the standard coordinate system.

Indeed, in the terminology of Markov processes, X is an additive functional of the Markov process Z.

The additivity for X implies certain measurability properties for the shifts:

3.4 LEMMA. *Let X be a Lévy process over the base $\mathscr{B}$. Let $\mathcal{G}$ be the filtration generated by it. For each t in $\mathbb{R}_+$, the mapping $\theta_t : \Omega \to \Omega$ is measurable with respect to $\mathcal{G}_{t+u}$ and $\mathcal{G}_u$ for every u in $\mathbb{R}_+$; in particular, θ_t is measurable with respect to $\mathcal{G}_\infty$ and $\mathcal{G}_\infty$.*

Proof. Fix t and u. For every s in $[0, u]$, we have $X_s \circ \theta_t = X_{t+s} - X_t$ by additivity, and $X_{t+s} - X_t$ is in $\mathcal{G}_{t+u}$. Since $\mathcal{G}_u$ is generated by X_s, with such s, this proves the first claim. The "particular" claim is obvious. □

Markov property

For a Lévy process, Markov property is the independence of future increments from the past at all times. The next theorem is the precise statement. Here, $\mathscr{B}$ is the stochastic base in 3.1, and $\mathcal{G}$ is the filtration generated by X. With the filtration $\mathcal{F}$ fixed, we write $\mathbb{E}_t$ for $\mathbb{E}(\cdot|\mathcal{F}_t)$, the conditional expectation operator given $\mathcal{F}_t$.

3.5 THEOREM. *Suppose that X is a Lévy process in $\mathbb{R}^d$ over the stochastic base $\mathscr{B}$. Then, for every time t, the process $X \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same law as X. Equivalently, for every bounded random variable V in $\mathcal{G}_\infty$,*

3.6 $$\mathbb{E}_t \, V \circ \theta_t \;=\; \mathbb{E}\, V, \quad t \in \mathbb{R}_+.$$

REMARK. The restriction to bounded V is for avoiding questions of existence for expectations. Of course, 3.6 extends to all positive V in $\mathcal{G}_\infty$ and to all integrable V in $\mathcal{G}_\infty$, and further.

Proof. a) We start by observing that the Lévy property 3.3c is equivalent to saying that, for every bounded Borel function f on $\mathbb{R}^d$,

3.7 $$\mathbb{E}_t \, f \circ X_u \circ \theta_t \;=\; \mathbb{E}\, f \circ X_u, \quad t, u \in \mathbb{R}_+.$$

b) We show next that 3.6 holds for V having the form $V = V_n$, where

3.8 $$V_n \;=\; f_1(X_{u_1})\, f_2(X_{u_2} - X_{u_1}) \cdots f_n(X_{u_n} - X_{u_{n-1}})$$

for some bounded Borel functions $f_1, \cdots, f_n$ on $\mathbb{R}^d$ and some times $0 < u_1 < u_2 < \cdots < u_n$.

The claim is true for $n = 1$ in view of 3.7. We make the induction hypothesis that the claim is true for n and consider it for $n+1$. Observe that, with $u = u_n$ and $v = u_{n+1} - u_n$ for simplicity of notation, we have

$$V_{n+1} \;=\; V_n \cdot W \circ \theta_u, \quad \text{where} \quad W \;=\; f_{n+1} \circ X_v.$$

Thus, writing $\mathbb{E}_t = \mathbb{E}_t\ \mathbb{E}_{t+u}$ and recalling that $\theta_u \circ \theta_t = \theta_{t+u}$, we get

$$\begin{aligned}\mathbb{E}_t\ V_{n+1}\circ\theta_t &= \mathbb{E}_t\ \mathbb{E}_{t+u}(V_n\circ\theta_t)(W\circ\theta_{t+u})\\ &= \mathbb{E}_t\ V_n\circ\theta_t\ \mathbb{E}_{t+u}\ W\circ\theta_{t+u}\\ &= \mathbb{E}_t\ V_n\circ\theta_t\ \mathbb{E}\ W \ = \ \mathbb{E}\ V_n\ \mathbb{E}\ W \ = \ \mathbb{E}\ V_{n+1}\end{aligned}$$

where we used 3.7 to justify the third equality sign, induction hypothesis to justify the fourth, and 3.7 again for the fifth. So 3.6 holds for every V of the form 3.8.

c) Borel functions f having the form $f(x_1, \cdots, x_n) = f_1(x_1) \cdots f_n(x_n)$ generate the Borel σ-algebra on $(\mathbb{R}^d)^n$. Thus, by the monotone class theorem, part (b) of the proof implies that 3.6 holds for every bounded V having the form

$$V = f(X_{u_1}, X_{u_2} - X_{u_1}, \cdots, X_{u_n} - X_{u_{n-1}}) \qquad \square$$

for some bounded Borel f and some times $0 < u_1 < \cdots < u_n$. Since the increments of X generate the σ-algebra $\mathcal{G}_\infty$, the proof is completed through another application of the monotone class theorem.

Strong Markov property

This is the analog of the Markov property where the deterministic time t is replaced by a stopping time T. The setup is the same, and we write $\mathbb{E}_T$ for the conditional expectation $\mathbb{E}(\cdot|\mathcal{F}_T)$. However, in formulating 3.6 with T, we face a problem: if $T(\omega) = \infty$, then $\theta_T\omega = \theta_{T(\omega)}\omega$ makes no sense and $X_\infty(\omega)$ is not defined. The following is to handle the problem.

3.9 Convention. *Suppose that $Z(\omega)$ is well-defined for every ω for which $T(\omega) < \infty$. Then, the notation $Z\ 1_{\{T<\infty\}}$ stands for the random variable that is equal to Z on $\{T < \infty\}$ and to 0 on $\{T = \infty\}$.*

The convention is without ambiguity. If Z is already defined for all ω, then $Z\ 1_{\{T<\infty\}}$ is equal to 0 on $\{T = \infty\}$ since $x \cdot 0 = 0$ for all x in $\bar{\mathbb{R}} = [-\infty, +\infty]$. With this convention, the following is the strong Markov property. Here $\bar{\mathcal{G}}_\infty$ is the completion of $\mathcal{G}_\infty$ in $\mathcal{H}$, that is $\bar{\mathcal{G}}_\infty = \mathcal{G}_\infty \vee \mathcal{N}$ where $\mathcal{N}$ is the σ-algebra generated by the collection of negligible events in $\mathcal{H}$.

3.10 Theorem. *Suppose that X is a Lévy process over the base $\mathscr{B}$. Let T be a stopping time of $\mathcal{F}$. Then, for every bounded random variable V in $\bar{\mathcal{G}}_\infty$,*

3.11 $$\mathbb{E}_T\ V\circ\theta_T\ 1_{\{T<\infty\}} \ = \ (\mathbb{E}\ V)\ 1_{\{T<\infty\}}.$$

Remark. If $T < \infty$ almost surely, then $1_{\{T<\infty\}}$ can be deleted on both sides. In words, the preceding theorem states the following: on the event $\{T < \infty\}$, the future process $X\circ\theta_T$ is independent of the past $\mathcal{F}_T$ and has the same law as X. On the event $\{T = \infty\}$, there is no future and nothing to be said.

Proof. Let X and T be as hypothesized. In view of the defining property for $\mathbb{E}_T$, it is sufficient to show that

3.12 $$\mathbb{E}\, 1_H\, V\circ\theta_T\, 1_{\{T<\infty\}} \;=\; \mathbb{E}\, 1_{H\cap\{T<\infty\}}\, \mathbb{E}\, V$$

for every H in $\mathcal{F}_T$ and bounded positive V in $\bar{\mathcal{G}}_\infty$. Moreover, for every V in $\bar{\mathcal{G}}_\infty$ there is V_0 in $\mathcal{G}_\infty$ such that $V = V_0$ almost surely, and it is enough to show 3.12 for V_0. Hence, it is enough to prove 3.12 for H in $\mathcal{F}_T$ and bounded positive V in $\mathcal{G}_\infty$. We do this in a series of steps.

a) Assume, further, that T is countably-valued. Let D be its range intersected with $\mathbb{R}_+$. Then, $\{T<\infty\}$ is equal to the union of $\{T=t\}$ over t in D, and $H\cap\{T=t\}\in\mathcal{F}_t$ for every t. Thus, starting with the monotone convergence theorem, the left side of 3.12 becomes

$$\sum_{t\in D}\mathbb{E}\, 1_{H\cap\{T=t\}} V\circ\theta_t \;=\; \sum_{t\in D}\mathbb{E}\, 1_{H\cap\{T=t\}}\, \mathbb{E}_t\, V\circ\theta_t \;=\; \mathbb{E}\, 1_{H\cap\{T<\infty\}}\, \mathbb{E}\, V,$$

where the last equality sign is justified by the Markov property that $\mathbb{E}_t V\circ\theta_t = \mathbb{E}\, V$. This proves 3.12 for T countably-valued.

b) Now we remove the restriction on T but assume that

3.13 $$V \;=\; f\circ X_u$$

for some $u>0$ and some bounded positive continuous function f on $\mathbb{R}^d$. Let (T_n) be the approximating sequence of stopping times discussed in Propositions V.1.20 and V.7.12: each T_n is countably-valued, $T_n<\infty$ on $\{T<\infty\}$, and the sequence decreases to T. Since $H\cap\{T<\infty\}\in\mathcal{F}_T\subset\mathcal{F}_{T_n}$, we get from 3.12 with T_n that

3.14 $$\mathbb{E}\, 1_{H\cap\{T<\infty\}}\, V\circ\theta_{T_n} \;=\; \mathbb{E}\, 1_{H\cap\{T<\infty\}}\, \mathbb{E}\, V.$$

On the event $\{T<\infty\}$, we have $T_n<\infty$ for every n, and

$$X_u\circ\theta_{T_n} \;=\; X_{T_n+u}-X_{T_n}\to X_{T+u}-X_T = X_u\circ\theta_T$$

almost surely, since (T_n) is decreasing to T and X is right-continuous. Since V has the form 3.13 with f continuous and bounded, it follows that $V\circ\theta_{T_n}\to V\circ\theta_T$ almost surely and, thus, the left side of 3.14 converges to the left side of 3.12 by the bounded convergence theorem. This proves 3.12 for V having the form 3.13 with f bounded, positive, and continuous.

c) Since continuous $f:\mathbb{R}^d\mapsto\mathbb{R}$ generate the Borel σ-algebra on $\mathbb{R}^d$, and since V satisfying 3.13 is a vector space and a monotone class, it follows that 3.12 and therefore 3.11 holds for V having the form 3.13 with f bounded, positive, Borel.

d) There remains to extend 3.11 to arbitrary bounded positive V in $\mathcal{G}_\infty$. This is done exactly as in the parts (b) and (c) of the proof of the Markov property, Theorem 3.5: put T wherever t appears and append the factor $1_{\{T<\infty\}}$ on each side of every equation having t in it.

Processes with bounded jumps

We have seen examples of Lévy processes X with $\mathbb{E}X_t$ equal to $+\infty$ (increasing stable processes) and also examples where $\mathbb{E}X_t$ does not exist (Cauchy processes). As an application of the strong Markov property, we now show that such anomalies are possible only if X has jumps of unbounded size. The converse is false: as gamma processes exemplify, X may have jumps of arbitrarily large size and still have finite moments of all orders.

3.15 Proposition. *Let X be a Lévy process in $\mathbb{R}^d$ over the base $\mathscr{B}$. Suppose that all its jumps are bounded in magnitude by some fixed constant. Then, for every t, the variable $|X_t|$ has finite moments of all orders.*

Proof. a) Fix a constant b in $(0, \infty)$. Suppose that all the jumps (if any) have magnitudes bounded by b. The claim is that, then, $\mathbb{E}\,|X_t|^k < \infty$ for every integer $k \geq 1$. To prove this, it is enough to show that the distribution of $|X_t|$ has an exponential tail; indeed, we shall show that there exists a constant c in $(0, 1)$ such that

3.16 $$\mathbb{P}\{\,|X_t| > (1+b)n\,\} \leq e^t\,c^n, \quad n \in \mathbb{N}.$$

b) Let R be the time of exit from the unit ball, that is,

$$R = \inf\{\,t > 0 \;:\; |X_t| > 1\,\}.$$

Note that $|X_R| \leq 1 + b$ since the worst that can happen is that X exits the unit ball by a jump of magnitude b. Moreover, since $X_0 = 0$ and X is right-continuous, $R > 0$ almost surely; hence,

3.17 $$c = \mathbb{E}\,e^{-R} < 1.$$

Finally, note that $R < \infty$ almost surely. This follows from the impossibility of containing the sequence (X_m) within the unit ball, since X_m is the sum of m independent and identically distributed variables.

c) Let T be a finite stopping time of $\mathcal{F}$ and consider

$$T + R\circ\theta_T = \inf\{t > T \;:\; |X_t - X_T| > 1\}.$$

By the strong Markov property at T, by Theorem 3.10 with $V = e^{-R}$,

3.18 $$\mathbb{E}\,e^{-(T+R\circ\theta_T)} = \mathbb{E}\,e^{-T}\,\mathbb{E}_T\,e^{-R}\circ\theta_T = \mathbb{E}\,e^{-T}\,\mathbb{E}\,e^{-R} = c\,\mathbb{E}\,e^{-T}.$$

d) Put $T_0 = 0$ and define $T_n, n \geq 1$, recursively by setting $T_{n+1} = T_n + R\circ\theta_{T_n}$. Since $R < \infty$ almost surely, so is $T_1 = R$ and so is $T_2 = T_1 + R\circ\theta_{T_1}$, and so on. Thus, T_n is an almost surely finite stopping time for each n, and using 3.18 repeatedly we get

3.19 $$\mathbb{E}\,e^{-T_n} = c^n, \quad n \in \mathbb{N}.$$

e) Finally, consider the bound 3.16. Note that T_{n+1} is the first time t after T_n such that $|X_t - X_{T_n}| > 1$. Thus, for fixed t and ω,

$$|X_t(\omega)| \;>\; (1+b)n \;\Rightarrow\; T_n(\omega) \;<\; t \;\Rightarrow\; e^{-T_n(\omega)} \;>\; e^{-t}.$$

Hence, the left side of 3.16 is less than or equal to

$$\mathbb{P}\{\, e^{-T_n} \;>\; e^{-t}\,\} \;\le\; e^t\, \mathbb{E}\, e^{-T_n} \;=\; e^t\, c^n\,.$$

by Markov's inequality and 3.19. This completes the proof. □

On the definitions

Consider Definitions 1.1 and 3.3. They differ at two points: the existence of shift operators and the conditions of right-continuity and augmentedness for the filtration $\mathcal{F}$. The shifts are for reasons of convenience and clarity. For instance, replacing $X_u \circ \theta_t$ with $X_{t+u} - X_t$ would eliminate the shifts in Definition 3.3 and in Theorem 3.5 on the Markov property. The same is true more generally; we can eliminate the shifts entirely, without loss of real content, but with some loss in brevity; for example, in Theorem 3.10 on the strong Markov property, we need to replace $X_u \circ \theta_T = X_{T+u} - X_T$ with $\hat{X}_u$ and, instead of $R\circ\theta_T$, we need to introduce $\hat{R}$, which is obtained from $\hat{X}$ by the same formula that obtains R from X. We use the shifts for the clarity and economy achieved through their use; see Exercise 3.24 to see that they can always be introduced without loss of generality.

The conditions on the filtration $\mathcal{F}$ are more serious. We illustrate the issue with an example. Suppose that X is a Wiener process and T is the time of hitting some fixed level $b > 0$, that is, $T \;=\; \inf\{t > 0 \;:\; X_t > b\}$. Since X is adapted to $\mathcal{F}$, and $\mathcal{F}$ is right-continuous, T is a stopping time of $\mathcal{F}$. If $\mathcal{F}$ were *not* right-continuous, T might fail to be a stopping time of it. For instance, T is not a stopping time of $\mathcal{G}$, the filtration generated by X ; this can be inferred from the failure of Galmarino's test, Exercise V.1.28.

Nevertheless, the conditions on $\mathcal{F}$ of Definition 3.3 are natural in addition to being advantageous. We show next that, starting with the filtration $\mathcal{G}$ generated by X, we can always use the augmentation $\bar{\mathcal{G}}$ as the filtration $\mathcal{F}$.

Augmenting the raw process

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, X a stochastic process with state space $\mathbb{R}^d$, and $\mathcal{G}$ the filtration generated by X. Let $(\Omega, \bar{\mathcal{H}}, \bar{\mathbb{P}})$ be the completion of $(\Omega, \mathcal{H}, \mathbb{P})$, and let $\mathcal{N}$ be the σ-algebra generated by the collection of negligible sets in $\bar{\mathcal{H}}$. We denote by $\bar{\mathcal{G}}$ the augmentation of $\mathcal{G}$ in $(\Omega, \bar{\mathcal{H}}, \bar{\mathbb{P}})$, that is, $\bar{\mathcal{G}}_t \;=\; \mathcal{G}_t \vee \mathcal{N}$, the σ-algebra generated by the union of $\mathcal{G}_t$ and $\mathcal{N}$. See Section 7 of Chapter V for these and for the notation $\mathcal{G}_{t+} = \cap_{\varepsilon>0}\ \mathcal{G}_{t+\varepsilon}$, and recall that the right-continuity for $\bar{\mathcal{G}}$ means that $\cap_{\varepsilon>0}\ \bar{\mathcal{G}}_{t+\varepsilon} = \bar{\mathcal{G}}_t$ for every t.

3.20 THEOREM. *Suppose that, over* $(\Omega, \mathcal{H}, \mathbb{P})$*, the process* X *is Lévy with respect to* $\mathcal{G}$ *in the sense of Definition 1.1. Then,*

a) over $(\Omega, \bar{\mathcal{H}}, \bar{\mathbb{P}})$ *, the process* X *is Lévy with respect to* $\bar{\mathcal{G}}$ *in the sense of Definition 1.1, and*

b) the filtration $\bar{\mathcal{G}}$ *is augmented and right-continuous.*

Before proving this, we note two interesting corollaries; one concerning $\mathcal{G}_{t+}$ and the other the special case $\bar{\mathcal{G}}_0$. Since $\mathcal{G}_{t+\varepsilon} \subset \bar{\mathcal{G}}_{t+\varepsilon}$, we have $\mathcal{G}_{t+} \subset \cap_{\varepsilon>0} \bar{\mathcal{G}}_{t+\varepsilon}$, and the last σ-algebra is equal to $\bar{\mathcal{G}}_t$ by the preceding theorem. Since $\bar{\mathcal{G}}_t = \mathcal{G}_t \vee \mathcal{N}$, we see that $\mathcal{G}_{t_+} \subset \mathcal{G}_t \vee \mathcal{N}$; in words, the extra wisdom gained by an infinitesimal peek into the future consists of events that are either negligible or almost sure. In particular, since $X_0 = 0$ almost surely, $\mathcal{G}_0 \subset \mathcal{N}$ and we obtain the following corollary, called *Blumenthal's zero-one law.*

3.21 COROLLARY. *Every event in* $\bar{\mathcal{G}}_0$ *has probability* 0 *or* 1.

Going back to arbitrary t, we express the finding $\mathcal{G}_{t+} \subset \bar{\mathcal{G}}_t$ in terms of random variables; recall that $\bar{\mathcal{G}}_t = \mathcal{G}_t \vee \mathcal{N}$, which means that every random variable in $\bar{\mathcal{G}}_t$ differs from one in $\mathcal{G}_t$ over a negligible set.

3.22 COROLLARY. *Fix* t *in* $\mathbb{R}_+$. *For every random variable* V *in* $\mathcal{G}_{t+}$ *there is a random variable* V_0 *in* $\mathcal{G}_t$ *such that* $V = V_0$ *almost surely.*

Proof of Theorem 3.20

a) We prove the first claim first. Suppose X is as hypothesized. Since the restriction of $\bar{\mathbb{P}}$ to $\mathcal{H}$ is equal to $\mathbb{P}$, the events in $\mathcal{H}$ that are $\mathbb{P}$-almost sure are also events in $\bar{\mathcal{H}}$ that are $\bar{\mathbb{P}}$-almost sure. Thus, the regularity 1.1a of X over $(\Omega, \mathcal{H}, \mathbb{P})$ remains as regularity over $(\Omega, \bar{\mathcal{H}}, \bar{\mathbb{P}})$.

For the Lévy property 1.1b, we first observe that X is such over $(\Omega, \mathcal{H}, \mathbb{P})$ with respect to $\mathcal{G}$ if and only if

3.23 $$\mathbb{E}\, V\, f\circ(X_{t+u} - X_t) = \mathbb{E}\, V\, \mathbb{E}\, f\circ X_u, \quad t, u \in \mathbb{R}_+,$$

for every bounded Borel f on $\mathbb{R}^d$ and bounded variable V in $\mathcal{G}_t$. Since $\bar{\mathbb{P}}$ coincides with $\mathbb{P}$ on $\mathcal{H}$, we may replace $\mathbb{E}$ with the expectation operator $\bar{\mathbb{E}}$ with respect to $\bar{\mathbb{P}}$. Finally, if $\bar{V}$ is a bounded variable in $\bar{\mathcal{G}}_t = \mathcal{G}_t \vee \mathcal{N}$, and f as above, there is V in $\mathcal{G}_t$ such that $\bar{V} = V$ almost surely (under $\bar{\mathbb{P}}$). Thus, in 3.23, we may replace $\mathbb{E}$ with $\bar{\mathbb{E}}$ and V with $\bar{V}$; the result is the Lévy property 1.1b over $(\Omega, \bar{\mathcal{H}}, \bar{\mathbb{P}})$ with respect to $\bar{\mathcal{G}}$.

b) We are working on the complete probability space $(\Omega, \bar{\mathcal{H}}, \bar{\mathbb{P}})$ to show that the augmentation $\bar{\mathcal{G}}$ is also right-continuous. We start at $t = 0$. Let (ε_n) be a sequence decreasing strictly to 0. For $n \geq 1$, let

$$\mathcal{H}_n = \sigma\{X_t - X_s : \quad \varepsilon_n \leq s < t \leq \varepsilon_{n-1}\}.$$

Since X is a Lévy process (shown in part (a)), the σ-algebras $\mathcal{H}_1, \mathcal{H}_2, \cdots$ are independent. By Theorem II.5.12, Kolmogorov's zero-one law, the tail

σ-algebra defined by $(\mathcal{H}_n)$ is trivial, that is, the tail σ-algebra is contained in $\mathcal{N}$. But, since $\mathcal{H}_{n+1} \vee \mathcal{H}_{n+2} \vee \cdots = \mathcal{G}_{\varepsilon_n}$, the tail σ-algebra is equal to

$$\cap_n \ \mathcal{G}_{\varepsilon_n} = \cap_{\varepsilon>0} \ \mathcal{G}_\varepsilon = \mathcal{G}_{0+}.$$

We have shown that $\mathcal{G}_{0+} \subset \mathcal{N}$. We use this to show that $\bar{\mathcal{G}}$ is right-continuous. Fix t ; let $\hat{\mathcal{G}}$ be the filtration generated by the process $\hat{X}$, where $\hat{X}_u = X_{t+u} - X_t$ for every time u. Since $\hat{X}$ is a Lévy process, what we have just shown applies to $\hat{\mathcal{G}}$ and we have $\hat{\mathcal{G}}_{0+} \subset \mathcal{N}$. It follows that

$$\bigcap_{\varepsilon>0} \bar{\mathcal{G}}_{t+\varepsilon} = \bigcap_{\varepsilon} (\mathcal{G}_t \vee \hat{\mathcal{G}}_\varepsilon \vee \mathcal{N}) = \mathcal{G}_t \vee \mathcal{N} \vee \hat{\mathcal{G}}_{0+} = \mathcal{G}_t \vee \mathcal{N} = \bar{\mathcal{G}}_t$$

because $\bar{\mathcal{G}}_{t+\varepsilon} = \mathcal{G}_{t+\varepsilon} \vee \mathcal{N}$ and $\mathcal{G}_{t+\varepsilon} = \mathcal{G}_t \vee \hat{\mathcal{G}}_\varepsilon$. In words, $\bar{\mathcal{G}}$ is right-continuous as claimed.

Exercises

3.24 *Processes of canonical type.* The aim is to introduce the probability law of a Lévy process in a concrete fashion. As a byproduct, this will show that every Lévy process is equivalent to a Lévy process of the type in Definition 3.3.
Setup.

i) Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, X a Lévy process with state space $\mathbb{R}^d$ in the sense of Definition 1.1, and $\mathcal{G}$ the filtration generated by X.

ii) Let W be the collection of all mappings $w : \mathbb{R}_+ \mapsto \mathbb{R}^d$ with $w(0) = 0$. Let Y_t be the coordinate mapping with $Y_t(w) = w(t)$, and let $\theta_t : W \mapsto W$ be defined by $\theta_t w(u) = w(t+u) - w(t)$. Define $\mathcal{K}_t$ to be the σ-algebra on W generated by $Y_s, s \le t$, and let $\mathcal{K} = (\mathcal{K}_t)$.

iii) Define the transformation $\varphi : \Omega \mapsto W$ by letting $\varphi\omega$ to be the path $X(\omega) : t \mapsto X_t(\omega)$.

a) Note that $Y_t \circ \varphi = X_t$ for every t. Use this to show that φ is measurable with respect to $\mathcal{G}_t$ and $\mathcal{K}_t$ for every t, and with respect to $\mathcal{G}_\infty$ and $\mathcal{K}_\infty$, and therefore with respect to $\mathcal{H}$ and $\mathcal{K}_\infty$.

b) It follows that $\mathbb{Q} = \mathbb{P} \circ \varphi^{-1}$ is a probability measure on $(W, \mathcal{K}_\infty)$. This $\mathbb{Q}$ is the distribution of X.

c) Show that, over the probability space $(W, \mathcal{K}_\infty, \mathbb{Q})$, the process $Y = (Y_t)$ is a Lévy process, in the sense of Definition 1.1, with respect to its own filtration $\mathcal{K}$.

d) Let $\bar{\mathcal{K}}$ be the augmentation of $\mathcal{K}$ in the completion $(W, \bar{\mathcal{K}}_\infty, \bar{\mathbb{Q}})$ of $(W, \mathcal{K}_\infty, \mathbb{Q})$. Show that

$$\mathscr{B} = (W, \bar{\mathcal{K}}_\infty, \bar{\mathcal{K}}, \theta, \bar{\mathbb{Q}})$$

is a stochastic base in the sense of Definition 3.1. Show that $Y = (Y_t)$ is a Lévy process over $\mathscr{B}$ in the sense of Definition 3.3.

4 Characterizations for Wiener and Poisson

Throughout this section $\mathscr{B}$ is the stochastic base introduced in Definition 3.1, and X is a Lévy process over $\mathscr{B}$ as in Definition 3.3. The aim is to characterize the three basic processes in qualitative terms: Poisson, the archetypical pure-jump process; Wiener, the continuous process *par excellence*; and compound Poisson process, whose paths are step functions.

Poisson processes

Recall that a Poisson process is a Lévy process whose increments are Poisson distributed. A number of characterizations were listed in Theorems VI.5.5 and VI.5.9. The following is a small addition.

4.1 Theorem. *The Lévy process X is Poisson if and only if it is a counting process.*

Proof. Sufficiency was shown in Theorem VI.5.9. To show the necessity, suppose that every increment of the Lévy process X has a Poisson distribution, $X_{t+u} - X_t$ with mean cu, where c is a constant in $\mathbb{R}_+$. Then, every increment takes values in $\mathbb{N}$ almost surely, which implies that almost every path is an increasing step function taking values in $\mathbb{N}$. To show that X is a counting process, there remains to show that every jump is of size 1 almost surely.

Fix t. Let H_t be the event that there is a jump of size 2 or more during $[0, t]$. Subdivide the interval $[0, t]$ into n intervals of equal length. The event H_t implies that, of the increments over those n subintervals, at least one increment is equal to 2 or more. Thus, by Boole's inequality,

$$\mathbb{P}\,(\,H_t\,) \;\leq\; n\;(1 - e^{-ct/n} - (ct/n)\;e^{-ct/n})$$

since each increment has the Poisson distribution with mean ct/n. Letting $n \to \infty$ we see that $\mathbb{P}\,(\,H_t\,) \,=\, 0$. Taking the union of H_t over $t = t_1, t_2, \cdots$ for some sequence (t_n) increasing to infinity, we see that, almost surely, no jump exceeds 1 in size. □

In the preceding proof, the sufficiency was by appealing to Theorem VI.5.9. The bare-hands proof of the latter theorem can be replaced with the following: Suppose that the Lévy process X is a counting process. Then, all jumps are bounded in size by 1, and Proposition 3.15 shows that $\mathbb{E}X_t < \infty$. Thus, the stationarity of the increments implies that $\mathbb{E}X_t = ct$ for some finite constant c, and the Lévy property implies that $M_t \;=\; X_t - ct$ defines a martingale M. Thus, by Theorem V.6.13, the process X is Poisson.

Wiener and continuous Lévy processes

According to the earlier definition, the Lévy process X is a Wiener process if it is continuous and X_t has the Gaussian distribution with mean 0 and variance t. The following shows that continuity is enough.

4.2 THEOREM. *Suppose that the Lévy process X is continuous and has state space $\mathbb{R}$. Then, it has the form*

$$X_t \;=\; bt + c\,W_t, \quad t \in \mathbb{R}_+,$$

where b and c are constants in $\mathbb{R}$, and W is a Wiener process over the base $\mathscr{B}$.

Proof. Assume X is such. Proposition 3.15 shows that X_t has finite mean and variance; there exist constants b and c such that $\mathbb{E}X_t = bt$ and $\mathrm{Var}X_t = c^2 t$. If $c = 0$, then there is nothing left to prove. Assuming that $c \neq 0$, define

$$W_t \;=\; (X_t - bt)/c, \quad t \in \mathbb{R}_+.$$

Then W is a continuous Lévy process over the base $\mathscr{B}$, and the Lévy property can be used to show that W is a continuous martingale, and so is $(W_t^2 - t)_{t\in\mathbb{R}_+}$, both with respect to the filtration $\mathcal{F}$. It follows from Proposition V.6.21 that W is Wiener. □

The preceding proof is via Proposition V.6.21, and the latter's proof is long and difficult. It is possible to give a direct proof using the Lévy property more fully: Start with W being a Lévy process with $\mathbb{E}\,W_t = 0$ and $\mathrm{Var}\,W_t = t$. For each integer $n \geq 1$,

$$W_1 \;=\; Y_{n,1} + \cdots + Y_{n,n}$$

where $Y_{n,j}$ is the increment of W over the interval from $(j-1)/n$ to j/n. Since W is Lévy, those increments are independent and identically distributed with mean 0 and variance $1/n$. Now the proof of the classical central limit theorem (III.8.1) can be adapted to show that W_1 has the standard Gaussian distribution. Thus, W is a Wiener process.

Continuous Lévy processes in $\mathbb{R}^d$

If $W^1, \ldots, W^d$ are independent Wiener processes, then $W = (W^1, \ldots, W^d)$ is a d-dimensional Wiener process. Obviously, it is a continuous Lévy process in $\mathbb{R}^d$. We now show that, conversely, every continuous Lévy process in $\mathbb{R}^d$ is obtained from such a Wiener process by a linear transformation plus some drift.

4.3 THEOREM. *Suppose that the Lévy process X in $\mathbb{R}^d$ is continuous. Then,*

$$X_t \;=\; bt + c\;W_t, \quad t \in \mathbb{R}_+,$$

for some vector b in $\mathbb{R}^d$, some $d \times d'$ matrix c, and a d'-dimensional Wiener process W over the base $\mathscr{B}$.

Remark. Let v be the covariance matrix for X_1. Then, d' is the rank of v (of course $d' \le d$), and $v = cc^T$ with c^T denoting the transpose of c.

Proof. Suppose X continuous. For r in $\mathbb{R}^d$, consider the linear combination $r \cdot X_t$ of the coordinates of X_t. The process $r \cdot X$ is a continuous Lévy process in $\mathbb{R}$. It follows from Theorem 4.2 that $r \cdot X_1$ has a one-dimensional Gaussian distribution, and this is true for every r in $\mathbb{R}^d$. Thus, X_1 has a d-dimensional Gaussian distribution with some mean vector b in $\mathbb{R}^d$ and some $d \times d$ matrix v of covariances.

The matrix v is symmetric and positive definite (that is, $v = v^T$ and $r \cdot vr \ge 0$ for every r in $\mathbb{R}^d$). Let d' be its rank. There exists some $d \times d'$ matrix c of rank d' such that $v = cc^T$, that is,

4.4 $$v_{ij} = \sum_{k=1}^{d'} c_{ik} c_{jk}, \quad i, j = 1, 2, \cdots, d.$$

We define a matrix a as follows. If $d' = d$, put $a = c$. If $d' < d$, the matrix c has exactly d' linearly independent rows, which we may assume are the rows $1, 2, \ldots, d'$ by re-labeling the coordinates of X ; we let a be the $d' \times d'$ matrix formed by those first d' rows of c. Obviously a is invertible; let $\hat{a}$ be its inverse. Define, for $i = 1, \ldots, d'$,

4.5 $$W_t^i = \sum_{k=1}^{d'} \hat{a}_{ik} \, (X_t^k - b_k t), \quad t \in \mathbb{R}_+.$$

It is clear that $W = (W^1, \ldots, W^{d'})$ is a continuous Lévy process in $\mathbb{R}^{d'}$, and W_t has the d'-dimensional Gaussian distribution with mean vector 0 and covariances

$$\begin{aligned}
\mathbb{E}\, W_t^i \, W_t^j &= \sum_{m=1}^{d'} \hat{a}_{im} \sum_{n=1}^{d'} \hat{a}_{jn} \, \mathbb{E}\, (X_t^m - b_m t)(X_t^n - b_n t) \\
&= \sum_{m=1}^{d'} \sum_{n=1}^{d'} \hat{a}_{im} \hat{a}_{jn} v_{mn} t \\
&= \sum_{m=1}^{d'} \sum_{n=1}^{d'} \hat{a}_{im} \hat{a}_{jn} \sum_{k=1}^{d'} a_{mk} a_{nk} t = \delta_{ij} t,
\end{aligned}$$

where the third equality follows from 4.4 since $c_{mk} = a_{mk}$ and $c_{nk} = a_{nk}$ for $m, n \le d'$. This shows that W is d'-dimensional Wiener. Reversing 4.4 and 4.5 shows that X is as claimed. □

Compound Poisson processes

We adopt the construction in Example 1.2c as the definition for compound Poisson processes. Several other constructions were mentioned previously in

this chapter and the last. The following characterization theorem summarizes the previous results and extends them onto the modern setting. This is basic.

4.6 THEOREM. *The Lévy process X over the base $\mathscr{B}$ is a compound Poisson process if and only if one (and therefore all) of the following statements holds.*

a) Almost every path of X is a step function.

b) There is a Poisson process (N_t) over $\mathscr{B}$ and, independent of it, an independency (Y_n) of identically distributed $\mathbb{R}^d$-valued variables such that

4.7
$$X_t = \sum_{n=1}^{\infty} Y_n \, 1_{\{n \le N_t\}}, \quad t \in \mathbb{R}_+.$$

c) There is a Poisson random measure M on $\mathbb{R}_+ \times \mathbb{R}^d$ whose mean has the form Leb $\times \lambda$ *with some finite measure λ on $\mathbb{R}^d$ such that*

4.8
$$X_t = \int_{[0,t]\times\mathbb{R}^d} M(ds, dx)\, x, \quad t \in \mathbb{R}_+.$$

4.9 REMARKS. a) The proof will show that the Poisson random measure M is adapted to the filtration $\mathcal{F}$ and is homogeneous relative to the shifts θ_t, that is, $\omega \mapsto M(\omega, A)$ is in $\mathcal{F}_t$ for every Borel subset A of $[0,t] \times \mathbb{R}^d$, and

$$M(\theta_t \omega, B) = M(\omega, B_t)$$

for every Borel subset B of $\mathbb{R}_+ \times \mathbb{R}^d$, where B_t consists of the points $(t+u, x)$ with (u, x) in B.

b) The connection between 4.7 and 4.8 is as follows. Let $c = \lambda(\mathbb{R}^d) < \infty$ and put μ for the distribution $(1/c)\,\lambda$. Finiteness of c implies that the atoms of M can be labeled as points (T_n, Y_n) so that $0 < T_1 < T_2 < \cdots$ almost surely. The T_n are the successive jump times of X and form the Poisson process N, and the Y_n are the variables appearing in 4.7. The jump rate for N is c, and the distribution common to Y_n is μ.

Proof. $(a) \Rightarrow (b)$. Assume (a). Let N_t be the number of jumps of X over $(0,t]$. Since X is adapted to $\mathcal{F}$, so is N ; since X is a step process, N is a counting process; and since $N_u \circ \theta_t$ is the number of jumps of $X \circ \theta_t$ over $(0,u]$, we have the additivity of N with respect to the shifts. Moreover, $X \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same law as X (this is the Markov property, Theorem 3.5); thus, $N_u \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same distribution as N_u. In summary, N is a Lévy process over the base $\mathscr{B}$ and is a counting process. By Theorem 4.1, it must be a Poisson process with some rate c.

Let Y_n be the size of the jump by X at T_n. Then, 4.7 is obvious, and there remains to show that (Y_n) is independent of (T_n) and is an independency of variables having the same distribution, say, μ on $\mathbb{R}^d$. We start by showing that $R = T_1$ and $Z = Y_1$ are independent and a bit more. The distribution of

Z is μ, and the distribution of R is exponential with parameter c ; the latter is because N is Poisson with rate c. For t in $\mathbb{R}_+$ and Borel subset B of $\mathbb{R}^d$, we now show that

$$\text{4.10} \qquad \mathbb{P}\{\ R > t,\ Z \in B\ \} \ = \ e^{-ct}\ \mu(B).$$

Note that, if $R(\omega) > t$, then $Z(\omega) = Z(\theta_t\omega)$. So, by the Markov property for X, Theorem 3.5, the left side of 4.10 is equal to

$$\begin{aligned} \mathbb{P}\{\ R > t,\ Z\circ\theta_t \in B\ \} &= \mathbb{E}\ 1_{\{R>t\}}\ \mathbb{E}_t\ 1_B\circ Z\circ\theta_t \\ &= \mathbb{E}\ 1_{\{R>t\}}\ \mathbb{E}\ 1_B\circ Z \ = \ e^{-ct}\ \mu(B) \end{aligned}$$

as claimed in 4.10. Next, for $n \geq 1$, we note that

$$T_{n+1} - T_n \ = \ R\circ\theta_{T_n}, \quad Y_{n+1} \ = \ Z\circ\theta_{T_n},$$

and use the strong Markov property proved in Theorem 3.10 at the almost surely finite stopping times T_n . We get that

$$\begin{aligned} \text{4.11}\ \mathbb{P}\{\ T_{n+1} - T_n > t\,,\ Y_{n+1} \in B \mid \mathcal{F}_{T_n}\ \} \\ = \mathbb{P}\{\ R\circ\theta_{T_n} > t\,,\ Z\circ\theta_{T_n} \in B \mid \mathcal{F}_{T_n}\ \} = \mathbb{P}\{\ R > t\,,\ Z \in B\ \}. \end{aligned}$$

Putting 4.10 and 4.11 together shows that the sequences (T_n) and (Y_n) are independent, and the Y_n are independent and have the distribution μ. This completes the proof that $(a) \Rightarrow (b)$.

$(b) \Rightarrow (c)$. Assume (b) and let (T_n) be the sequence of successive jump times of N. It follows from Corollary VI.3.5 that the pairs (T_n, Y_n) form a Poisson random measure M on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean $c\text{Leb} \times \mu = \text{Leb} \times \lambda$, where $\lambda = c\mu$ is a finite measure on $\mathbb{R}^d$. It is obvious that, then, 4.7 and 4.8 are the same equation served up in differing notations.

$(c) \Rightarrow (a)$. Assume (c) ; then 4.8 shows that X is a Lévy process; and the paths are almost surely step functions, because the measure λ is finite. □

The best qualitative definition for compound Poisson processes is that they are Lévy processes whose paths are step functions. The following provides another equivalent condition for it. As before, X is a Lévy process in $\mathbb{R}^d$ over the base $\mathscr{B}$. We leave its proof to Exercise 4.14.

4.12 Proposition. *Almost every path of X is a step function if and only if the probability is strictly positive that*

$$R \ = \ \inf\{\ t > 0\ :\ X_t \neq 0\ \}$$

is strictly positive. Moreover, then $0 < R < \infty$ almost surely and has the exponential distribution with some parameter c in $(0, \infty)$.

4.13 REMARK. Obviously, R is a stopping time of $(\mathcal{G}_{t+})$. Thus, the event $\{R > 0\}$ belongs to $\mathcal{G}_{0+}$ and, by Blumenthal's zero-one law (Corollary 3.21), its probability is either 0 or 1. In other words, either $R = 0$ almost surely or $R > 0$ almost surely; in the former case the point 0 of $\mathbb{R}^d$ is said to be *instantaneous*, and in the latter case, *holding*. The preceding can now be restated: X is a compound Poisson process if and only if the point 0 is a holding point, in which case the holding time has an exponential distribution.

Exercise

4.14 *Proof of Proposition* 4.12. If the paths are step functions, then R is the time of first jump and is necessarily strictly positive. The following are steps leading to the sufficiency part, assuming that $R > 0$ almost surely in view of Remark 4.13.

a) Show that, if $R(\omega) > t$, then $R(\omega) = t + R(\theta_t\omega)$. Use the Markov property to show that the function $f(t) = \mathbb{P}\{R > t\}, t \in \mathbb{R}_+$, satisfies $f(t+u) = f(t)f(u)$.

b) Note that f is right-continuous and bounded; the case $f = 1$ is excluded by the standing assumption that X is not degenerate; show that the case $f = 0$ is also excluded. Thus, $f(t) = e^{-ct}$ for some constant c in $(0, \infty)$; in other words, $0 < R < \infty$ almost surely and R is exponential.

c) Show that, on the event $H = \{X_{R-} = X_R\}$, we have $X_{R-} = X_R = 0$ and $R \circ \theta_R = 0$. Use the strong Markov property to show that

$$\mathbb{P}(H) = \mathbb{P}(\, H \cap \{R \circ \theta_R = 0\}\,) = \mathbb{P}(H)\mathbb{P}\{R = 0\} = 0.$$

Hence, R is a jump time almost surely, obviously the first.

d) Define $T_1 = R$, and recursively put $T_{n+1} = T_n + R \circ \theta_{T_n}$. Show that the T_n form a Poisson process. Conclude that X is a step process.

5 Itô-Lévy Decomposition

This is to show the exact converse to Theorem 1.29: every Lévy process has the form given there. Throughout, $\mathscr{B} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$ is a stochastic base, and X is a stochastic process over it with state space $\mathbb{R}^d$. A random measure M on $\mathbb{R}_+ \times \mathbb{R}^d$ is said to be *Poisson over* $\mathscr{B}$ with Lévy measure λ if

5.1 a) $M(A)$ is in $\mathcal{F}_t$ for every Borel subset A of $[0, t] \times \mathbb{R}^d$,

b) $M(\theta_t\omega, B) = M(\omega, B_t)$ for every ω and t and Borel subset B of $\mathbb{R}_+ \times \mathbb{R}^d$, where $B_t = \{(t+u, x) : (u, x) \in B\}$, and

c) M is Poisson with mean $\text{Leb} \times \lambda$, and λ is a Lévy measure, that is, $\lambda\{0\} = 0$ and

$$\int_{\mathbb{R}^d} \lambda(dx)\ (|x|^2 \wedge 1) < \infty.$$

With this preparation, we list the main theorem of this section next. It is called the *Itô-Lévy decomposition theorem.* Its sufficiency part is Theorem 1.29. The necessity part will be proved in a series of propositions of interesting technical merit. Recall Notation 1.27 and its meaning.

5.2 THEOREM. *The process X is a Lévy process over $\mathscr{B}$ if and only if, for every t in $\mathbb{R}_+$,*

$$X_t \;=\; bt + cW_t + \int_{[0,t]\times\mathbb{B}} [M(ds,dx) - ds\lambda(dx)]x + \int_{[0,t]\times\mathbb{B}^c} M(ds,dx)x$$

for some vector b in $\mathbb{R}^d$, some $d\times d'$ matrix c, some d'-dimensional Wiener process W over $\mathscr{B}$, and, independent of W, a random measure M on $\mathbb{R}_+\times\mathbb{R}^d$ that is Poisson over $\mathscr{B}$ with some Lévy measure λ.

5.3 COROLLARY. *If X is a Lévy process, then its characteristic exponent is*

$$\psi(r) \;=\; ib\cdot r - \tfrac{1}{2}r\cdot vr + \int_{\mathbb{R}^d} \lambda(dx)\;(e^{ir\cdot x} - 1 - ir\cdot x\;1_{\mathbb{B}}(x)), \quad r\in\mathbb{R}^d$$

for some b in $\mathbb{R}^d$, some $d\times d$ matrix v that is symmetric and positive definite, and some measure λ on $\mathbb{R}^d$ that is a Lévy measure. Conversely, if (b,v,λ) is such a triplet, there is a Lévy process whose characteristic exponent is ψ above.

The preceding corollary is immediate from Theorems 5.2 and 1.29. This is basically the *Lévy-Khinchine formula* stated in stochastic terms. Obviously, b and λ are as in Theorem 5.2, and $v = cc^T$.

5.4 REMARKS. a) *Characteristic triplet.* This refers to (b,v,λ) ; it defines the law of X by defining the characteristic exponent in the canonical form given in the preceding corollary.

b) *Semimartingaleness.* It follows from the theorem above that every Lévy process is a semimartingale; see Remark 1.34c also.

c) *Special cases.* Characterizations for Lévy processes with special properties can be deduced from the theorem above and discussion in Section 1. See Theorems 4.2 and 4.3 for X continuous, Theorem 4.6 for X step process, Theorem 1.12 for X pure-jump.

d) *Increasing processes.* Suppose that the state space is $\mathbb{R}$ and X is increasing. Then, in the theorem above and its corollary, we must have $v = 0, \lambda(-\infty,0] = 0$, and, further, λ must satisfy 1.13. It is usual to represent such X in the form

$$X_t \;=\; at + \int_{[0,t]\times(0,\infty)} M(ds,dx)\;x, \quad t\in\mathbb{R}_+.$$

The corresponding characteristic triplet is $(b,0,\lambda)$ with λ as noted and

$$b \;=\; a + \int_{(0,1]} \lambda(dx)x.$$

Jumps exceeding ε in magnitude

Recall the notation $\Delta X_t = X_t - X_{t-}$ and also Remark 1.9 on the sparseness of jumps exceeding $\varepsilon > 0$ in magnitude.

5.5 PROPOSITION. *Suppose that X is a Lévy process over $\mathscr{B}$. For $\varepsilon > 0$, let*

$$X_t^\varepsilon = \sum_{s \le t} \Delta X_s \, 1_{\{|\Delta X_s| > \varepsilon\}}, \quad t \in \mathbb{R}_+.$$

Then, X^ε is a compound Poisson process over $\mathscr{B}$.

Proof. a) It is clear that X^ε is adapted to $\mathcal{F}$ and is additive with respect to shifts. Since X_u^ε is $\mathcal{G}_\infty$-measurable, it follows from the Markov property (Theorem 3.5) that $X_u^\varepsilon \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same distribution as X_u^ε. Thus, X^ε is a Lévy process over the base $\mathscr{B}$.

b) Since the paths of X are right-continuous and left-limited with $X_0 = 0$, the paths of X^ε are step functions; see Remark 1.9. It follows from the characterization theorem 4.6 that X^ε is a compound Poisson process. □

Some independence

This is to show that X^ε above is independent of $X - X^\varepsilon$. Generally, proofs of independence are easy consequences of assumptions made beforehand. This is a rare case where the proof requires serious work.

5.6 PROPOSITION. *Suppose that X is a Lévy process over $\mathscr{B}$. For fixed $\varepsilon > 0$, let X^ε be as defined in the preceding proposition. Then, X^ε and $X - X^\varepsilon$ are independent Lévy processes over $\mathscr{B}$.*

Proof. a) The preceding proposition has shown that X^ε is a Lévy process over $\mathscr{B}$ and more; part (*a*) of its proof is easily adapted to show that $X - X^\varepsilon$ is a Lévy process over $\mathscr{B}$, and that the pair $(X^\varepsilon, X - X^\varepsilon)$ is a Lévy process over $\mathscr{B}$ with state space $\mathbb{R}^d \times \mathbb{R}^d$. To show that X^ε and $X - X^\varepsilon$ are independent, then, is reduced to showing that X_t^ε and $X_t - X_t^\varepsilon$ are independent for every t. To show the latter, it is enough to show that, for every q and r in $\mathbb{R}^d$,

5.7 $$\mathbb{E} \exp\left[iq\cdot X_t^\varepsilon + ir\cdot(X_t - X_t^\varepsilon) \right] = \left(\mathbb{E} \exp iq\cdot X_t^\varepsilon\right)\left(\mathbb{E} \exp ir\cdot(X_t - X_t^\varepsilon)\right).$$

b) Fix q and r in $\mathbb{R}^d$. Recall (see 1.3) that the characteristic functions on the right side of 5.7 have the form $e^{t\varphi}$ and $e^{t\psi}$ for some complex numbers φ and ψ depending on q and r respectively. With these notations, the Lévy property for X^ε and, separately, for $X - X^\varepsilon$ shows that

$$L_t = 1 - \exp\left(iq \cdot X_t^\varepsilon - t\varphi\right), \quad M_t = 1 - \exp\left(ir \cdot (X_t - X_t^\varepsilon) - t\psi\right)$$

define complex-valued $\mathcal{F}$-martingales. We shall show that

5.8 $$\mathbb{E}\, L_t M_t = 0, \quad t \in \mathbb{R}_+,$$

thus showing 5.7 and completing the proof.

c) Fix $t > 0$. Fix $n \geq 1$. Let $\mathcal{D}$ be the subdivision of $(0,t]$ into n equi-length intervals of the type $(,]$. For each interval A in $\mathcal{D}$, if $A = (u,v]$, we put $L_A = L_v - L_u$ and $M_A = M_v - M_u$. With this notation, since $L_0 = M_0 = 0$,

$$L_t M_t \;=\; \sum_{A\in\mathcal{D}} L_A \sum_{B\in\mathcal{D}} M_B .$$

Take expectations on both sides. Note that $\mathbb{E}\, L_A M_B = 0$ if A and B are disjoint; this is by the martingale property for L and M. It follows that

5.9 $$\mathbb{E}\, L_t M_t \;=\; \mathbb{E}\, R_n, \quad \text{where} \quad R_n \;=\; \sum_{A\in\mathcal{D}} L_A M_A .$$

d) We show now that $|R_n| \leq R$ where R is an integrable random variable free of n. First, observe that

5.10 $$|M_A| \;\leq\; 2 \sup_{s\leq t} |M_s|, \quad A \in \mathcal{D},$$

and that, by Doob's norm inequality (V.3.26),

5.11 $$\mathbb{E} \sup_{s\leq t} |M_s|^2 \;\leq\; 4\, \mathbb{E}\, |M_t|^2 \;<\infty .$$

Next, by the definition of L, with $X_A^\varepsilon = X_v^\varepsilon - X_u^\varepsilon$ for $A = (u,v]$,

$$\begin{aligned} |L_A| &= |e^{\,iq\cdot X_u^\varepsilon}\, e^{\,-u\varphi} - e^{\,iq\cdot X_v^\varepsilon}\, e^{\,-v\varphi}| \\ &= |e^{\,-u\varphi} - e^{\,-v\varphi} + e^{\,-v\varphi}\,(1 - e^{\,iq\cdot X_A^\varepsilon})| \\ &\leq |e^{\,-u\varphi} - e^{\,-v\varphi}| + |e^{\,-v\varphi}| \cdot |1 - e^{\,iq\cdot X_A^\varepsilon}| \\ &\leq \int_A a e^{as}\, ds \;+\; 2\, e^{at}\, 1_{\{X_A^\varepsilon \neq 0\}} \end{aligned}$$

where we put $a = |\varphi|$ and observed that $|1 - e^{\,iq\cdot x}|$ is equal to 0 if $x = 0$ and is bounded by 2 if $x \neq 0$. Thus,

5.12 $$\sum_{A\in\mathcal{D}} |L_A| \;\leq\; e^{at} \;+\; 2\, e^{at} \sum_{A\in\mathcal{D}} 1_{\{X_A^\varepsilon\neq 0\}} \;\leq\; e^{at}\,(1 + 2\,K_t)$$

where K_t is the number of jumps X^ε has during $(0,t]$. Since K_t has the Poisson distribution with some mean ct, and therefore variance ct,

5.13 $$\mathbb{E}\,(1 + 2K_t)^2 \;<\; \infty .$$

It follows from 5.10 and 5.12 that

$$|R_n| \;\leq\; 2 \sup_{s\leq t} |M_s| \sum_{A\in\mathcal{D}} |L_A| \;\leq\; 2\, e^{at}\, (\sup_{s\leq t} |M_s|)(1 + 2K_t) \;=\; R$$

where R is integrable in view of 5.11 and 5.13.

e) Finally, we let $n \to \infty$ in 5.9. Since X^ε is a step process, the martingale L has finitely many jumps during $[0,t]$ and is smooth between the jumps. Thus,

$$\lim R_n = \sum_{s \le t} (L_s - L_{s-})(M_s - M_{s-}) = 0,$$

where the sum is over the finitely many jump times s of X^ε and the last equality is because $M_s - M_{s-} = 0$ at those times s since $X - X^\varepsilon$ has no jumps in common with X^ε. In view of part (d) above, the dominated convergence theorem applies, and we have

$$\lim \mathbb{E}\, R_n = 0.$$

This proves 5.8 via 5.9 and completes the proof. □

Jump measure

This is to show that jumps of X are governed by a Poisson random measure. We start by defining the random measure, to be called the *jump measure* of X. Recall the notation ΔX_t and also the set D_ω of discontinuities of the path $X(\omega)$. Let $\omega \in \Omega$, and A Borel subset of $\mathbb{R}_+ \times \mathbb{R}^d$; if the path $X(\omega)$ is right-continuous, left-limited, and $X_0(\omega) = 0$, we put

5.14 $$M(\omega, A) = \sum_{t \in D_\omega} 1_A(t, \Delta X_t(\omega));$$

for all other ω, put $M(\omega, A) = 0$. For each ω, this defines a counting measure on $\mathbb{R}_+ \times \mathbb{R}^d$.

5.15 Proposition. *Suppose that X is a Lévy process over $\mathscr{B}$. Then, M is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean* Leb $\times \lambda$, *where λ is a Lévy measure on $\mathbb{R}^d$; that is, 5.1 holds.*

Proof. It follows from the definition of M and augmentedness of $\mathcal{F}$ that the condition 5.1a holds. Similarly, 5.1b is satisfied by the additivity of X. There remains to show 5.1c.

For $\varepsilon > 0$, let M_ε be the trace of M on $\mathbb{R}_+ \times \varepsilon\mathbb{B}^c$, where $\varepsilon\mathbb{B}^c$ is the set of all εx with x outside the unit ball $\mathbb{B}$. Comparing 5.14 with the definition of X^ε in Proposition 5.5, we see that M_ε is the jump measure of X^ε. Since X^ε is compound Poisson, it follows from Theorem 4.6 that M_ε is a Poisson random measure with mean $\mu_\varepsilon = \text{Leb} \times \lambda_\varepsilon$, where λ_ε is a finite measure on $\mathbb{R}^d$.

It is obvious that λ_ε puts all its mass outside $\varepsilon\mathbb{B}$. Define the measure λ on $\mathbb{R}^d$ by letting, for g positive Borel, λg be the increasing limit of $\lambda_\varepsilon g$ as $\varepsilon > 0$ decreases to 0. Put $\mu = \text{Leb} \times \lambda$. It is obvious that $\lambda\{0\} = 0$, and that μ_ε is the trace of μ on $\mathbb{R}_+ \times \varepsilon\mathbb{B}^c$.

Let f be a positive Borel function on $\mathbb{R}_+ \times \mathbb{R}^d$. Then, $\omega \mapsto M_\varepsilon f(\omega)$ is a random variable for each ε, and $M_\varepsilon f(\omega)$ increases to $Mf(\omega)$ as $\varepsilon \to 0$. Thus, Mf is a random variable, and

$$\mathbb{E}\, e^{-Mf} = \lim_{\varepsilon\downarrow 0} \mathbb{E}\, e^{-M_\varepsilon f} = \lim_{\varepsilon\downarrow 0} \exp_- \mu_\varepsilon(1-e^{-f}) = \exp_- \mu(1-e^{-f}).$$

Thus, M is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean $\mu = \text{Leb} \times \lambda$. The proof is complete via the next lemma. □

5.16 Lemma. *The measure λ is a Lévy measure.*

Proof. We have noted that $\lambda\{0\} = 0$ by definition. Recall that λ_ε is the Lévy measure for the compound Poisson process X^ε. In particular, then, λ_ε is finite for $\varepsilon = 1$. Thus, to show that the λ-integral of $x \mapsto |x|^2 \wedge 1$ is finite, it is sufficient to show that

5.17 $$\int_{\mathbb{B}} \lambda(dx)\, |x|^2 < \infty.$$

By Proposition 5.6 above, X^ε and $X - X^\varepsilon$ are independent. Thus,

$$|\mathbb{E}\, e^{ir\cdot X_t}| = |\mathbb{E}\, e^{ir\cdot(X_t - X_t^\varepsilon)}\, \mathbb{E}\, e^{ir\cdot X_t^\varepsilon}| \le |\mathbb{E}\, e^{ir\cdot X_t^\varepsilon}|$$
$$\le |\exp\, t\int_{\varepsilon\,\mathbb{B}^c} \lambda(dx)\, (e^{ir\cdot x} - 1)|$$

5.18 $$\le \exp_-\, t\int_{\varepsilon\,\mathbb{B}^c} \lambda(dx)\, (1 - \cos r\cdot x) \le \exp_-\, \frac{t}{4}\int_{\mathbb{B}\setminus\varepsilon\mathbb{B}} \lambda(dx)\, |r\cdot x|^2$$

for r in $\mathbb{B}$, where the last step used the observation that $\varepsilon\,\mathbb{B}^c = \mathbb{R}^d\setminus\varepsilon\,\mathbb{B} \supset \mathbb{B}\setminus\varepsilon\mathbb{B}$ and $1 - \cos u \ge u^2/4$ for $|u| \le 1$. Since the left-most member is free of ε, we let $\varepsilon \to 0$ in the right-most member to conclude that

$$\int_{\mathbb{B}} \lambda(dx)\, |r\cdot x|^2 < \infty$$

for every r in $\mathbb{B}$; this implies 5.17 as needed. □

Proof of the decomposition theorem 5.2

Suppose that X is a Lévy process. Recall Proposition 5.15 about the jump measure M. In terms of it, $X^e = X^\varepsilon$ with $\varepsilon = 1$ is given by

5.19 $$X_t^e = \int_{[0,t]\times\mathbb{B}^c} M(ds, dx)\, x, \quad t \in \mathbb{R}_+.$$

Since λ is a Lévy measure, Theorem 1.23 (and using Notation 1.27) yields a Lévy process X^d over $\mathscr{B}$ through

5.20 $$X_t^d = \int_{[0,t]\times\mathbb{B}} [\, M(ds, dx) - ds\, \lambda(dx)\,]\, x, \quad t \in \mathbb{R}_+.$$

Poisson nature of M implies that X^d and X^e are independent. The definition of M shows that $X - X^d - X^e$ has almost surely continuous paths, and the Markov property for X shows that the latter is a Lévy process over $\mathscr{B}$. Thus, by Theorem 4.2,

5.21 $$X_t - X_t^d - X_t^e \;=\; bt \;+\; c\,W_t, \quad t \in \mathbb{R},$$

where b, c, W are as claimed in Theorem 5.2. Putting 5.19, 5.20, and 5.21 together yields the decomposition wanted, and the main claim of the theorem is proved, except for the independence of W and M.

For $\varepsilon > 0$, the process X^ε of 5.5 is determined by M_ε, the trace of M on $\mathbb{R}_+ \times \varepsilon\mathbb{B}^c$, whereas W is determined by $X - X^\varepsilon$. Independence of X^ε and $X - X^\varepsilon$ proved in Proposition 5.6 implies that W and M_ε are independent. This is true for every $\varepsilon > 0$, and $M_\varepsilon f$ increases to Mf for every positive Borel f on $\mathbb{R}_+ \times \mathbb{R}^d$. It follows that W and M are independent.

6 Subordination

This is about time changes using increasing Lévy processes as clocks. In deterministic terms, the operation is as follows. Imagine a clock, something like the odometer of a car; suppose that, when the clock points to the number t, the standard time is s_t. Imagine, also, a particle whose position in $\mathbb{R}^d$ is z_s when the standard time is s. Then, when the clock shows t, the particle's position is z_{s_t}.

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. Let $S = (S_t)$ be an increasing process. Let $Z = (Z_s)$ be a process with state space $\mathbb{R}^d$. Define

6.1 $$X_t(\omega) \;=\; Z_{S_t(\omega)}(\omega), \qquad \omega \in \Omega,\; t \in \mathbb{R}_+.$$

Then, $X = (X_t)$ is said to be obtained by *subordinating* Z to S, and S is called the *subordinator*. We write $X = Z_S$ to express 6.1.

The concept of subordination can be extended: all that is needed is that the subordinator's state space be contained in the parameter set of Z. For instance, the compound Poisson process X of Example 1.2c is obtained by subordinating

$$Z_n = Z_0 + Y_1 + \cdots + Y_n, \qquad n \in \mathbb{N},\; Z_0 = 0,$$

to the Poisson process $N = (N_t)$; this is immediate upon noting $X_t = Z_{N_t}$ is another way of expressing the sum defining X_t in 1.2c. Another such example is where $Z = (Z_n)_{n\in\mathbb{N}}$ is a Markov chain with some state space $(E, \mathcal{E})$, and $N = (N_t)$ is a Poisson process independent of Z; then, $X = Z_N$ is a Markov process (in continuous time) with state space $(E, \mathcal{E})$.

For the remainder of this section, Z and S will be independent Lévy processes. To keep the setting simple, and also because there are three processes and two time scales, we use Definition 1.1 for Lévy processes (with respect to their own filtrations).

Main results

6.2 THEOREM. *Let S be an increasing Lévy process. Let Z be a Lévy process in $\mathbb{R}^d$. Suppose that the two are independent. Then, $X = Z_S$ is a Lévy process in $\mathbb{R}^d$.*

Proof. Since S is increasing and the regularity condition 1.1a holds for S and for Z, the same condition holds for X. We now show that 1.1b holds for X with $\mathcal{F}$ as the filtration generated by X.

Fix times $0 \le t_0 < t_1 < \cdots < t_n < \infty$ and let $f_1, \cdots, f_n$ be positive Borel functions on $\mathbb{R}^d$. Conditioning on the σ-algebra $\mathcal{G}_\infty$ generated by S, using the independence of Z and S, and also the Lévy property for Z, we obtain

$$\mathbb{E} \prod_{i=1}^{n} f_i(\,X_{t_i} - X_{t_{i-1}}\,) \;=\; \mathbb{E} \prod_{i=1}^{n} g_i(\,S_{t_i} - S_{t_{i-1}}\,),$$

where $g_i(s) = \mathbb{E}\; f_i \circ Z_s$. Since S is a Lévy process, the right side is equal to

$$\prod_{i=1}^{n} \mathbb{E}\; g_i(\,S_{t_i} - S_{t_{i-1}}\,) \;=\; \prod_{i=1}^{n} \mathbb{E}\; g_i(S_{t_i - t_{i-1}}).$$

But, by the definition of g_i and the independence of S and Z,

$$\mathbb{E}\; g_i(S_t) \;=\; \mathbb{E}\; f_i(Z_{S_t}) \;=\; \mathbb{E}\; f_i(X_t).$$

We have shown that the increments of X over the intervals $(t_{i-1}, t_i]$, $1 \le i \le n$, are independent and stationary. □

In the remainder of this section, we present a number of examples of subordination and give a characterization of the law of X in terms of the laws of S and Z. For the present, we list the following useful result without proof; it is a corollary to Theorem 6.18 to be proved at the end of the section.

6.3 PROPOSITION. *Let Z, S, X be as in Theorem 6.2. Suppose that S is pure-jump with Lévy measure ν. Then, the Lévy measure of X is,*

$$\lambda(B) \;=\; \int_{(0,\infty)} \nu(ds)\; \mathbb{P}\{\; Z_s \in B\backslash\{0\}\;\}, \qquad \textit{Borel } B \subset \mathbb{R}^d. \tag{6.4}$$

The heuristic reasoning behind this is as follows. Since S is an increasing pure-jump process, S_t is equal to the sum of the lengths of the intervals $(S_{u-}, S_u]$, $u \le t$. This implies that X_t is equal to the sum of the increments of Z over those intervals. Now, $\nu(ds)$ is the rate (per unit of clock time) of S-jumps of size belonging to the small interval ds around the value s; and, given that an interval $(S_{u-}, S_u]$ has length s, the corresponding increment of Z has the same distribution as Z_s. See Theorem 6.18 for more.

Gamma subordinators

6.5 EXAMPLE. *Wiener subordinated to gamma.* In Theorem 6.2, take Z to be a Wiener process, and let S be a gamma process with shape rate a and scale parameter c; see Example 1.21. Then, given S_t, the conditional distribution of X_t is Gaussian with mean 0 and variance S_t; thus,

$$\mathbb{E}\ \exp\ irX_t \ =\ \mathbb{E}\exp_{-}\frac{1}{2}r^2 S_t \ =\ \left(\frac{c}{c+r^2/2}\right)^{at} \ =\ \left(\frac{2c}{2c+r^2}\right)^{at}.$$

Hence, in the terminology of Example 1.21, X is a symmetric gamma process with shape rate a and scale parameter $\sqrt{2c}$. As described there, X is the difference of two independent gamma processes, each with rate a and scale $\sqrt{2c}$, say $X = X^+ - X^-$. Indeed, by the reasoning following Proposition 6.3,

$$X_t^+ \ =\ \sum_{u\le t} (Z_{S_u} - Z_{S_{u-}})^+, \qquad X_t^- \ =\ \sum_{u\le t} (Z_{S_u} - Z_{S_{u-}})^-,$$

each sum being over the countable set of jump times u of S.

The Lévy measure of X is as given in 1.21 with c there replaced by $\sqrt{2c}$ here. We re-derive it to illustrate 6.4: for x in $\mathbb{R}$,

$$\begin{aligned}\lambda(dx) &= dx \int_0^\infty ds\ a\ \frac{e^{-cs}}{s}\cdot\frac{e^{-x^2/2s}}{\sqrt{2\pi s}} \\ &= dx\ a\ \frac{1}{|x|}\int_0^\infty ds\ e^{-cs}\ \frac{|x|\,e^{-x^2/2s}}{\sqrt{2\pi s^3}} \ =\ dx\ a\ \frac{e^{-|x|\sqrt{2c}}}{|x|},\end{aligned}$$

where we evaluated the last integral by recognizing it as the Laplace transform of a stable distribution with index $\frac{1}{2}$; see VI.4.10. See also Exercise 6.26 for the d-dimensional version of this example.

Stable subordinators

Subordination operation is especially interesting when the subordinator S is an increasing stable process with index a; the index must be in $(0,1)$ since S is increasing. Exercise 6.29 is an example where Z is a gamma process. The following is about the case when Z is stable; it shows that the stability of Z is inherited by X.

6.6 PROPOSITION. *Let S, Z, X be as in Theorem 6.2. Suppose that S is an increasing a-stable process, $a\in(0,1)$, and that Z is a b-stable process in $\mathbb{R}^d$, $b\in(0,2]$. Then, X is a stable process in $\mathbb{R}^d$ with index ab.*

6.7 REMARK. In particular, taking Z to be a Wiener process in $\mathbb{R}^d$, the subordination yields an isotropic stable process in $\mathbb{R}^d$ with index $ab = 2a$. Every isotropic stable process X with index in $(0,2)$ is obtained in this manner by taking a such that $2a$ is equal to the index of X.

Proof. Let S, Z, X be as assumed. Since X is Lévy by Theorem 6.2, we need to check only its stability; we need to show that X_t has the same distribution as $t^{1/ab}X_1$. Fix t. Since S is a-stable, S_t has the same distribution as $t^{1/a}S_1$, which implies that X_t has the same distribution as Z_{uS_1} with $u = t^{1/a}$. On the other hand, since Z is b-stable, Z_{us} has the same distribution as $u^{1/b}Z_s$. Thus, since S and Z are independent, Z_{uS_1} has the same distribution as $u^{1/b}Z_{S_1} = t^{1/ab}X_1$. □

6.8 Example. *Cauchy in* $\mathbb{R}^d$. This is to illustrate the uses of the preceding theorem; we shall re-establish the results on Cauchy processes in $\mathbb{R}^d$. Let Z be a Wiener process in $\mathbb{R}^d$. Independent of it, let S be the increasing stable process of Example 2.1 with $a = 1/2$ and $c = 1/\sqrt{2\pi}$; then, the Lévy measure is $\nu(ds) = ds\, 1/\sqrt{2\pi s^3}$, and (see 2.1 and VI.4.10)

$$\text{6.9} \qquad \mathbb{E}\, e^{-pS_t} = e^{-t\sqrt{2p}}, \qquad \mathbb{P}\{\, S_t \in ds \,\} = ds\, \frac{te^{-t^2/2s}}{\sqrt{2\pi s^3}},$$

for p and s positive. According to the preceding proposition, X is an isotropic stable process with index $ab = \frac{1}{2}\cdot 2 = 1$, a Cauchy process.

It follows from the well-remembered formula $\mathbb{E}\, \exp\, ir\cdot Z_t = \exp_-\, t|r|^2/2$ and the independence of S and Z that

$$\text{6.10} \qquad \mathbb{E}\, e^{ir\cdot X_t} = \mathbb{E}\, \exp_-\, \frac{1}{2} S_t\, |r|^2 = e^{-t|r|}, \qquad r \in \mathbb{R}^d,$$

in view of the Laplace transform in 6.9. So, X is the standard Cauchy process in $\mathbb{R}^d$. The distribution of X_t can be obtained by inverting the Fourier transform in 6.10; we do it directly from the known distributions of S_t and Z_s:

$$\begin{aligned} \mathbb{P}\{\, X_t \in dx \,\} &= dx \int_0^\infty \mathbb{P}\{S_t \in ds\} \frac{e^{-|x|^2/2s}}{(2\pi s)^{d/2}} \\ \text{6.11} \qquad &= dx\, \frac{t\,\Gamma\left(\frac{d+1}{2}\right)}{[\,\pi t^2 + \pi |x|^2\,]^{(d+1)/2}} \end{aligned}$$

here we used 6.9, replaced $2s$ with $1/u$, and noted that the integral is a constant times a gamma density.

Comparing 6.10 with 2.27, we see that the Lévy measure of X is the measure λ given by 2.24 with $c = 2\Gamma\left(\frac{d+1}{2}\right)\Big/\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{1}{2}\right)$, and σ the uniform distribution on the unit sphere. Here is a confirmation of it in Cartesian coordinates: using Proposition 6.3,

$$\text{6.12} \quad \lambda(dx) = dx \int_0^\infty ds\, \frac{1}{\sqrt{2\pi s^3}} \cdot \frac{e^{-|x|^2/2s}}{(2\pi s)^{d/2}} = dx\, \frac{\hat{c}}{|x|^{d+1}}, \quad x \in \mathbb{R}^d,$$

where $\hat{c} = \Gamma\left(\frac{d+1}{2}\right)\Big/\pi^{(d+1)/2}$. For $d = 1$, this reduces to 2.15 as it should.

6.13 REMARK. The preceding exercise contains the distributions of S_t, X_t, and Z_t, namely, the strictly stable distributions with indices 1/2, 1, and 2. These three seem to be the only stable distributions that can be displayed explicitly in terms of common functions.

Transformation of laws under subordination

This is to characterize the probability law of X in terms of the laws of Z and S. To specify the laws of Z and X, we employ characteristic triplets. In general, for an arbitrary Lévy process X, we shall use the shorthand $X \sim (b, v, \lambda)$ to mean that X has (b, v, λ) as its characteristic triplet. We recall Corollary 5.3 and Remark 5.4a:

6.14 $$X \sim (b, v, \lambda) \ \Leftrightarrow\ \mathbb{E}\, e^{ir\cdot X_t} \ = \ \exp\ t\,[\,ib\cdot r - \frac{1}{2} r\cdot vr + \lambda f_r\,]$$

where $f_r\,(x) = e^{ir\cdot x} - 1 - ir\cdot x 1_{\mathbb{B}}(x)$. The following lemma is obvious and needs no proof.

6.15 LEMMA. *a) If X is a compound Poisson process with Lévy measure λ, then $X \sim (\lambda h, 0, \lambda)$, where $h(x) = x1_{\mathbb{B}}(x)$ for x in $\mathbb{R}^d$.*

b) If $X \sim (b, v, \lambda)$ and $X'_t = X_{at}$ for some fixed a in $\mathbb{R}_+$, then $X' \sim (ab, av, a\lambda)$.

c) If $X' \sim (b', v', \lambda')$ and $X'' \sim (b'', v'', \lambda'')$, and if X' and X'' are independent, then $X' + X'' \sim (b' + b'',\ v' + v'',\ \lambda' + \lambda'')$.

The next theorem gives a complete characterization of the law of $X = Z_S$. In preparation, we introduce

6.16 $$K(s, B) \ = \ \mathbb{P}\{\ Z_s \in B\,,\ Z_s \neq 0\ \}\,, \qquad s \in \mathbb{R}_+\,, \quad \text{Borel } B \subset \mathbb{R}^d$$

and note that K is a sub-probability transition kernel. For S we use the representation

6.17 $$S_t = at + S^o_t, \qquad t \in \mathbb{R}_+,$$

with a in $\mathbb{R}_+$ and S^o pure-jump with Lévy measure ν. This is the general form of an increasing Lévy process (see Remark 5.4d).

6.18 THEOREM. *Let Z, S, X be as in Theorem 6.2. Suppose that $Z \sim (b, v, \lambda)$. Then, with $h(x) = x1_{\mathbb{B}}(x)$ for x in $\mathbb{R}_d$,*

$$X \ \sim \ (\ ab + \nu K h,\ av,\ a\lambda + \nu K\ ).$$

Proof. a) We write $X_t = Z(S_t)$ for ease of notation. The process X is Lévy by Theorem 6.2; thus, the claim here is about the characteristic function of X_t. It follows from 6.17 that, for fixed t,

$$X_t \ = \ Z(at) + [\,Z(at + S^o_t) - Z(at)\,] \ = \ X'_t + X''_t,$$

say, where X_t' and X_t'' are independent, the former has the same distribution as $Z(at)$, and the latter as $Z(S_t^o)$. By part (*b*) of the last lemma, $X' \sim (ab, av, a\lambda)$; and by part (*c*), the triplet for $X_t = X_t' + X_t''$ is the sum of the triplets for X_t' and X_t''. Hence, the proof is reduced to showing that

6.19 $$X^o \;=\; Z(S^o) \;\sim\; (\;\nu K h,\; 0,\; \nu K\;).$$

b) Let S^ε be the pure-jump process where jumps are those of S with sizes exceeding $\varepsilon > 0$. Then, S^ε is a compound Poisson process, and its Lévy measure ν_ε is the trace of ν on (ε, ∞). Its successive jump times T_n form a Poisson process with rate $\nu\,(\varepsilon,\infty)$, and the corresponding sequence (U_n) of jump sizes is independent of (T_n) and is an independency with the distribution $\mu \;=\; \nu_\varepsilon \,/\, \nu(\varepsilon,\infty)$ for each U_n. It follows from this picture that

$$X_t^\varepsilon \;=\; Z(S_t^\varepsilon) \;=\; \sum_n \; Y_n\, 1_{\{T_n \le t\}},$$

where, with $U_0 = 0$,

$$Y_n \;=\; Z(U_0 + \cdots + U_n) \;-\; Z(U_0 + \cdots + U_{n-1}), \quad n \ge 1.$$

Note that (Y_n) is independent of (T_n) and is an independency with the common distribution

$$\mathbb{P}\{\; Y_1 \in B \;\} \;=\; \int_{\mathbb{R}_+} \mu(ds)\; \mathbb{P}\{\; Z_s \in B \;\}, \qquad \textit{Borel } B \subset \mathbb{R}^d.$$

Hence, X^ε is a compound Poisson process with Lévy measure $\nu_\varepsilon K$; and we were careful to exclude the mass at the origin which the distribution of Z_s might have. So, the characteristic exponent of X^ε is

6.20 $$\psi_\varepsilon(r) \;=\; \int_{(\varepsilon,\infty)} \nu(ds) \int_{\mathbb{R}^d} K(s,dx)\; (e^{ir\cdot x} - 1), \qquad r \in \mathbb{R}^d.$$

c) Let $\varepsilon \to 0$. Since S^o is pure-jump, S_t^ε increases to S_t^o, which implies that $Z(S_t^\varepsilon) \to Z(S_t^o -)$ by the left-limitedness of Z. But, for fixed s, we have $Z_s = Z_{s-}$ almost surely, and this remains true for $s = S_t^o$ by the independence of Z from S. Hence, $X_t^\varepsilon \to X_t^o$ almost surely, and the characteristic exponent of X^o is

6.21 $$\psi_o(r) \;=\; \lim_{\varepsilon \downarrow 0} \psi_\varepsilon(r).$$

d) Let φ be the characteristic exponent of Z. For $s \le 1$,

$$\left| \int_{\mathbb{R}^d} K(s,dx)\; (e^{ir\cdot x} - 1) \right| \;=\; \left| e^{s\varphi(r)} - 1 \right| \le s\; |\varphi(r)|\;;$$

and

$$\int_{(0,1]} \nu(ds)\; s < \infty$$

since ν is the Lévy measure of an increasing process. Thus, the dominated convergence theorem applies, and

$$\lim_{\varepsilon\downarrow 0}\int_{(\varepsilon,1]}\nu(ds)\int_{\mathbb{R}^d}K(s,dx)\,(e^{ir\cdot x}-1) \;=\; \int_{(0,1]}\nu(ds)\int_{\mathbb{R}^d}K(s,dx)\,(e^{ir\cdot x}-1).$$

Putting this together with 6.20 and 6.21, we get

$$6.22 \qquad \psi_o(r) \;=\; \int_{(0,\infty)}\nu(ds)\int_{\mathbb{R}^d}K(s,dx)\,(e^{ir\cdot x}-1).$$

e) We now show that, as 6.22 suggests, the Lévy measure of X^o is νK. Let M^ε be the jump measure of the process X^ε. We have shown in part (*b*) that it is Poisson with mean $Leb\times\nu_\varepsilon K$. For positive Borel functions f on $\mathbb{R}_+\times\mathbb{R}^d$, it is clear that $M^\varepsilon f$ increases to some limit $M^o f$ as $\varepsilon\to 0$, and since

$$\mathbb{E}e^{-M^\varepsilon f} \;=\; \exp_-\int_{\mathbb{R}_+}dt\int_{(\varepsilon,\infty)}\nu(ds)\int_{\mathbb{R}^d}K(s,dx)\,[1-e^{-f(t,x)}],$$

we have

$$\mathbb{E}e^{-Mf} \;=\; \exp_-\int_{\mathbb{R}_+}dt\int_{(0,\infty)}\nu(ds)\int_{\mathbb{R}^d}K(s,dx)\,[1-e^{-f(t,x)}]$$

by the monotone convergence theorem. Thus M^o is Poisson with mean $Leb\times\nu K$. It now follows from part (*c*) of the proof that M^o is the jump measure of X^o. Hence, in particular, the Lévy measure of X^o is νK.

f) Since νK is a Lévy measure on $\mathbb{R}^d$,

$$\psi_1(r)=\int_{\mathbb{B}}\nu K(dx)\,(e^{ir\cdot x}-1-ir\cdot x),$$

$$6.23 \qquad \psi_2(r)=\int_{\mathbb{B}^c}\nu K(dx)\,(e^{ir\cdot x}-1)$$

are well-defined complex numbers for each r in $\mathbb{R}^d$. Writing

$$ir\cdot x \;=\; (e^{ir\cdot x}-1)-(e^{ir\cdot x}-1-ir\cdot x)$$

and recalling that $h(x)=x1_{\mathbb{B}}(x)$, we see from 6.22 and 6.23 that

$$\left|\int_{\mathbb{R}^d}\nu K(dx)\;r\cdot h(x)\right| \;=\; \left|\psi_o(r)-\psi_2(r)-\psi_1(r)\right| \;<\infty.$$

Taking $r=e_j$, the jth unit vector, for $j=1,\cdots,d$, we see that νKh is a well-defined vector in $\mathbb{R}^d$, and that

$$6.24 \qquad \psi_o(r) \;=\; i\,(\nu Kh)\cdot r+\psi_1(r)+\psi_2(r).$$

In view of 6.23, this implies through 6.14 that $X^o\sim(\nu Kh,0,\nu K)$; hence, 6.19 is true, and the proof is complete. □

Exercises

6.25 *Symmetric gamma.* Let k_a be as defined in Exercise 1.49 , that is, k_a is the density function for the difference of two independent gamma variables with the same shape index a and the same scale parameter 1. Let X be as in Example 6.5.

a) Show that the density function for X_t is $\sqrt{2c}\, k_{at}(\sqrt{2c}\; x)$.
b) Show that

$$k_a(x) \;=\; \int_0^\infty du\; \frac{e^{-u}\, u^{a-1}}{\Gamma(a)} \cdot \frac{e^{-x^2/4u}}{\sqrt{4\pi u}}, \qquad x \in \mathbb{R}.$$

This is more appealing than its close relative, the *modified Bessel function* K_ν. The latter is given by

$$K_\nu(x) \;=\; \frac{1}{2}\left(\frac{x}{2}\right)^{-\nu} \int_0^\infty du\; e^{-u}\, u^{\nu-1} e^{-x^2/4u}, \quad \nu \in \mathbb{R}, \quad x \in \mathbb{R}_+.$$

Thus, for $a > 0$ and x in $\mathbb{R}$,

$$k_a(x) \;=\; \frac{|x/2|^{a-1/2}}{\sqrt{\pi}\,\Gamma(a)}\; K_{a-1/2}(\,|x|\,).$$

6.26 *Wiener subordinated to gamma.* In Theorem 6.2, let Z be a Wiener process in $\mathbb{R}^d$, and S a gamma process with shape rate a and scale parameter c . In view of Example 6.5 , every component of $X = Z_S$ is a symmetric gamma process with shape rate a and scale parameter $\sqrt{2c}$. The process X is isotropic.

a) Show that

$$\mathbb{E}\; e^{ir\cdot X_t} \;=\; \left(\frac{2c}{2c + |r|^2}\right)^{at}, \quad r \in \mathbb{R}^d.$$

b) Let λ be the Lévy measure of X. In spherical coordinates (see 1.52) , its spherical part σ is the uniform distribution on the unit sphere in $\mathbb{R}^d$, and its radial part ρ is given by

$$\rho(dv) \;=\; dv \int_0^\infty ds\; \frac{ae^{-cs}}{s} \cdot \frac{2v^{d-1}\, e^{-v^2/2s}}{(2s)^{d/2}\,\Gamma(d/2)}, \quad v > 0.$$

Show this. Show that

$$\rho(dv) \;=\; dv \cdot \frac{\Gamma(\frac{d+1}{2})\,\Gamma(\frac{1}{2})}{\Gamma(\frac{d}{2})} \cdot \frac{4a}{v}\, k_b(\sqrt{2c}\, v)$$

with $b = \frac{d+1}{2}$; see the preceding exercise for k_b .

6.27 *Stable subordinated to gamma.* Let Z be an isotropic a-stable process in $\mathbb{R}^d$ having the characteristic exponent $\psi(r) = -|r|^a$. Let S be a gamma process with shape rate b and scale parameter c. As usual, we assume that Z and S are independent. For $X = Z_S$, show that

$$\mathbb{E}\, e^{ir\cdot X_t} = \left(\frac{c}{c+|r|^a}\right)^{bt}, \quad r \in \mathbb{R}^d.$$

The distribution of X_t is called *Linnik distribution.*

6.28 *Continuation.* Suppose, now, that Z is an increasing stable process with index a, necessarily, $a \in (0,1)$. Then, $X = Z_S$ is an increasing pure-jump process. Suppose that the scale parameter is chosen to make Z_t have the Laplace transform e^{-tp^a}, $p \in \mathbb{R}_+$.

a) Show that

$$\mathbb{E}\, e^{-pX_t} = \left(\frac{c}{c+p^a}\right)^{bt}, \quad p \in \mathbb{R}_+.$$

b) Show that, when $a = 1/2$, the Lévy measure for X is

$$\lambda(dx) = dx\, b\, \frac{e^{c^2 x}}{x} \int_{c\sqrt{2x}}^{\infty} du\, \frac{e^{-u^2/2}}{\sqrt{2\pi}}, \quad x > 0.$$

6.29 *Gamma subordinated to stable.* Let Z be a gamma process with shape rate b and scale parameter 1. Let S be an increasing stable process with shape index a and scale c, that is, $\mathbb{E}\, \exp_- pS_t = \exp_- tcp^a$ for p in $\mathbb{R}_+$; here $c > 0$ and $a \in (0,1)$. Show that, then,

$$\mathbb{E}\, e^{-pX_t} = \exp_-\, tc\,[\,b\,\log(1+p)\,]^a, \quad p \in \mathbb{R}_+.$$

7 Increasing Lévy Processes

These processes play an important role in the theories of regeneration and Markov processes in continuous time. Moreover, they are useful as subordinators and interesting in themselves. In this section, we give a highly selective survey concentrating on potentials and hitting times.

Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space, $\mathcal{F} = (\mathcal{F}_t)$ is an augmented right-continuous filtration, and $S = (S_t)$ is an increasing Lévy process relative to $\mathcal{F}$. The assumptions on $\mathcal{F}$ are without loss of generality in view of Theorem 3.20.

We let b denote the drift rate and λ the Lévy measure for S. Thus, b is a constant in $\mathbb{R}_+$, and the measure λ on $\mathbb{R}_+$ satisfies 1.13 and $\lambda\{0\} = 0$. More explicitly,

7.1 $$S_t = bt + \int_{(0,t]\times\mathbb{R}_+} M(ds, dx)\, x, \quad t \in \mathbb{R}_+,$$

where M is Poisson on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $Leb \times \lambda$. We let π_t be the distribution of S_t and recall that, for p in $\mathbb{R}_+$,

7.2 $$\mathbb{E}e^{-p\,S_t} = \int_{\mathbb{R}_+} \pi_t(dx)\, e^{-px} = \exp_\, t\,[\,bp + \int_{\mathbb{R}_+} \lambda(dx)\, (1 - e^{-px})\,].$$

We exclude from further consideration the trivial case where $\lambda = 0$. When $b = 0$ and λ finite, S is a compound Poisson process, and its paths are step functions. Otherwise, S is strictly increasing.

Potential measure

For Borel subsets B of $\mathbb{R}_+$, we define

7.3 $$U(B) = \mathbb{E} \int_{\mathbb{R}_+} dt\; 1_B \circ S_t = \int_{\mathbb{R}_+} dt\; \pi_t(B),$$

the expected amount of time spent in B by S. Then, U is called the *potential measure* of S. Explicit computations are rare, but the Laplace transform

7.4 $$\hat{u}_p = \int_{\mathbb{R}_+} U(dx)\, e^{-px} = \int_{\mathbb{R}_+} dt\; \mathbb{E}\, e^{-pS_t}$$

is readily available: in view of 7.2,

7.5 $$[\,bp + \int_{\mathbb{R}_+} \lambda(dx)\, (1 - e^{-px})\,]\; \hat{u}_p = 1, \qquad p \in (0, \infty).$$

7.6 Example. a) *Poisson process.* Suppose that S is a Poisson process with rate c. Then, it spends an exponential amount with mean $1/c$ at each positive integer n. So, $U = (1/c)\,(\delta_0 + \delta_1 + \cdots)$, where δ_x is Dirac at x as usual.

b) *Stable process.* Suppose that S is increasing stable with index a; the index is necessarily in $(0, 1)$. Then, the Lévy measure has the density c/x^{a+1} with respect to Lebesgue on $(0, \infty)$; see 2.1. Choosing $c = a/\Gamma(1 - a)$, the Laplace transform for S_t becomes $\exp_\, tp^a$, and hence

$$\int_{\mathbb{R}_+} U(dx)\, e^{-px} = \frac{1}{p^a} = \int_{\mathbb{R}_+} dx\, \frac{e^{-px}\, x^{a-1}}{\Gamma(a)}.$$

It follows that the potential measure is absolutely continuous, and

$$U(dx) = dx\, \frac{x^{a-1}}{\Gamma(a)}, \quad x \in \mathbb{R}_+.$$

7.7 Remark. a) The measure U is diffuse, except when S is compound Poisson: if S is not compound Poisson, then it is strictly increasing, which

implies that the amount of time spent in the singleton $\{x\}$ is equal to zero. When S is compound Poisson, the point 0 is an atom for U, because S spends an exponential amount of time at 0 with parameter $c = \lambda(\mathbb{R}_+) < \infty$; there are atoms beyond 0 only if λ has atoms.

b) The potential measure is finite on compacts: For $B \subset [0, x]$,

$$U(B) \le \int_{\mathbb{R}_+} dt\ \mathbb{P}\{\ S_t \le x\ \} = \int_{\mathbb{R}_+} dt\ \mathbb{P}\{\ e^{-S_t} \ge e^{-x}\ \}$$
$$\le \int_{\mathbb{R}_+} dt\ e^x\ \mathbb{E}\ e^{-S_t}\ =\ e^x\ \hat{u}_1,$$

where the inequality is Markov's; and $\hat{u}_1 < \infty$ since the first factor on the left side of 7.5 is strictly positive for $p = 1$.

Absolute continuity of the potential

This is a closer examination of the equation 7.5 for the Laplace transform $\hat{u}_p$. We start by introducing a measure φ on $\mathbb{R}_+$:

7.8 $$\varphi(dx)\ =\ b\,\delta_0(dx) + dx\ \lambda(x, \infty)\ 1_{(0,\infty)}(x), \qquad x \in \mathbb{R}_+.$$

Since the Lévy measure λ satisfies 1.13, its tail $x \mapsto \lambda(x, \infty)$ is a real-valued decreasing locally integrable function on $(0, \infty)$. Thus, the measure φ is finite on compacts. Note that its Laplace transform is

$$\hat{\varphi}_p\ =\ \int_{\mathbb{R}_+} \varphi(dx)\ e^{-px}\ =\ b + \frac{1}{p} \int_{\mathbb{R}_+} \lambda(dx)\ (1 - e^{-px}), \qquad p > 0.$$

Hence, we may re-write 7.5 as $\hat{\varphi}_p\, \hat{u}_p = \frac{1}{p}$; in other words, the convolution of the measures U and φ is equal to the Lebesgue measure on $\mathbb{R}_+$, that is,

$$\int_{\mathbb{R}_+} \varphi(dx) \int_{\mathbb{R}_+} U(dy)\ 1_B(x + y)\ =\ \text{Leb}\ B,$$

or equivalently,

7.9 $$b\,U(B) + \int_B dx \int_{[0,x]} U(dy)\ \lambda(x - y, \infty)\ =\ \text{Leb}\ B, \qquad B \in \mathcal{B}_{\mathbb{R}_+}.$$

7.10 Remark. Suppose that $b = 0$. Then, the preceding equation shows that

7.11 $$\int_{[0,x]} U(dy)\ \lambda(x - y, \infty)\ =\ 1$$

for Lebesgue-almost every x in $(0, \infty)$. It is known that, in fact, this is true for every x in $(0, \infty)$. We shall show this when S is compound Poisson; see 7.25ff. The proof in the remaining case, where $b = 0$ and $\lambda(\mathbb{R}_+) = +\infty$, is famously difficult; see notes and comments for this chapter.

7.12 THEOREM. *Suppose that* $b > 0$. *Then,* U *is absolutely continuous and admits a bounded continuous function* $u : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *as its density; and*

7.13 $$b\,u(x) + \int_0^x dy\ u(y)\ \lambda(x-y,\infty) \;=\; 1, \qquad x \in \mathbb{R}_+.$$

Proof. It follows from 7.9 that $bU \le Leb$, which implies via Radon-Nikodym theorem that $U(dx) = dx\ u(x)$, $x \in \mathbb{R}_+$,for some positive Borel function bounded by $1/b$. Then, 7.9 implies that 7.13 holds for Leb-almost every x. Since u is bounded and $x \mapsto \lambda(x,\infty)$ is right-continuous and locally integrable, the second term on the left side of 7.13 is continuous in x. Thus, we may take u continuous, and 7.13 holds for every x. □

Level crossings

Let T_x denote the time of hitting (x,∞) by S; we call it also the time of crossing the level x:

7.14 $$T_x \;=\; \inf\{\, t \ge 0 \;:\; S_t > x \,\}, \qquad x \in \mathbb{R}_+.$$

Each T_x is a stopping time of $(\mathcal{G}_{t+})$, where $\mathcal{G}$ is the filtration generated by S; it is also a stopping time of $\mathcal{F}$ since $\mathcal{F}$ is right-continuous. The processes (S_t) and (T_x) are functional inverses of each other. If S is compound Poisson, then $T_0 > 0$ almost surely, and (T_x) is a step process. Otherwise, S is strictly increasing, and $T_0 = 0$ almost surely, and (T_x) is continuous.

7.15 PROPOSITION. *For fixed* x *and* t *in* $\mathbb{R}_+$,

7.16 $$\mathbb{P}\{\, T_x \le t \,\} \;=\; \begin{cases} \mathbb{P}\{\, S_t > x \,\} & \text{if } S \text{ is compound Poisson,} \\ \mathbb{P}\{\, S_t \ge x \,\} & \text{otherwise.} \end{cases}$$

In both cases,

$$\mathbb{E}\,T_x \;=\; U\,[0,x] \;=\; \int_0^\infty dt\ \mathbb{P}\{\, S_t \le x \,\}.$$

Proof. Pick ω such that the regularity conditions hold for the corresponding path of S. If S is compound Poisson, the path is a step function, and $T_x(\omega) \le t$ if and only if $S_t(\omega) > x$; this proves 7.16 in this case. Otherwise, the path is strictly increasing; then, $S_t(\omega) \ge x$ implies that $T_x \le t$, and the latter implies that $x \le S(\omega, T_x(\omega)) \le S(\omega,t)$; thus, $T_x(\omega) \le t$ if and only if $S_t(\omega) \ge x$, and this proves 7.16 in this case.

As to expected values, it follows from 7.16 that

$$\mathbb{E}\,T_x \;=\; \int_{\mathbb{R}_+} dt\ \mathbb{P}\{\, T_x > t \,\} \;=\; \begin{cases} U\,[0,x] & \text{if S is compound Poisson,} \\ U\,[0,x) & \text{otherwise .} \end{cases}$$

But if S is not compound Poisson, then U is diffuse (see Remark 7.7a), and we have $U\,[0,x) = U\,[0,x]$. □

Jumping across

In the remainder of this section we shall consider the joint distribution of T_x and the values of S just before and after T_x. We introduce (G for *gauche*, and D for *droit*)

7.17 $$G_x = S_{T_x-}, \quad D_x = S_{T_x}, \qquad x \in \mathbb{R}_+,$$

with the convention that $S_{0-} = 0$ always; see Figure 10 below. In general, $G_x \le x \le D_x$. Crossing into (x, ∞) occur either by drifting across x, which is the case on the event $\{ G_x = x = D_x \}$, or by jumping across x, which is the case on $\{ G_x < D_x \}$. The following gives the joint distribution in the jump case.

7.18 THEOREM. *Let $x \in (0,\infty)$. Let $f : (\mathbb{R}_+)^3 \mapsto \mathbb{R}_+$ be Borel. Then,*

$$\mathbb{E} f(T_x, G_x, D_x) 1_{\{ G_x \neq D_x \}} = \int_{\mathbb{R}_+} dt \int_{[0,x]} \pi_t(dy) \int_{[x-y,\infty)} \lambda(dz)\, f(t, y, y+z).$$
7.19

7.20 REMARK. Case of $x = 0$. If S is not compound Poisson, there is nothing to do, since $T_0 = 0$ and $G_0 = D_0 = 0$. If S is compound Poisson, then 7.19 remains true for $x = 0$: Then, T_0 is the time of first jump, which has the exponential distribution with parameter $c = \lambda(\mathbb{R}_+)$; and $G_0 = 0$ almost surely; and D_0 is the size of the first jump, which is independent of T_0 and has the distribution $(1/c)\,\lambda$; whereas, $\pi_t\{0\} = e^{-ct}$.

Proof. Fix $x > 0$ and f Borel. Let Z denote the random variable on the left side of 7.19. Being increasing, S can cross the level x only once. For almost every ω, therefore, there is at most one jump time t with $S_{t-}(\omega) \le x \le S_t(\omega)$; and if t is such, putting $z = S_t(\omega) - S_{t-}(\omega) > 0$, we obtain an

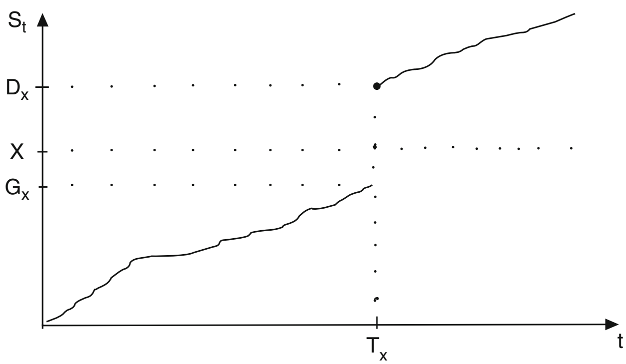


Figure 10: Level x is crossed at time T_x by a jump from the point G_x in $[0, x]$ to the point D_x in (x, ∞).

atom (t, z) of the measure $M(\omega, \cdot)$ defining $S(\omega)$; see 7.1. Thus,

$$Z = \int_{\mathbb{R}_+ \times (0,\infty)} M(dt, dz)\ f(t, S_{t-}, S_{t-} + z)\ 1_{\{S_{t-} \le x \le S_{t-} + z\}}\ ;$$

indeed, for almost every ω, the integral is a sum with at most one term, namely, the term corresponding to $t = T_x(\omega)$ if $S_t(\omega) - S_{t-}(\omega) = z > 0$. So, Z is a Poisson integral, and the integrand is predictable (see Theorem VI.6.2) since $t \mapsto S_{t-}$ is left-continuous and adapted. Hence,

$$\begin{aligned}\mathbb{E}\ Z = \mathbb{E} \int_{\mathbb{R}_+} dt \int_{(0,\infty)} \lambda(dz)\ f(t, S_{t-}, S_{t-} + z)\ 1_{\{S_{t-} \le x \le S_{t-} + z\}} \\ = \mathbb{E} \int_{\mathbb{R}_+} dt \int_{(0,\infty)} \lambda(dz)\ f(t, S_t, S_t + z)\ 1_{\{S_t \le x \le S_t + z\}},\end{aligned}$$

where the last equality is justified by noting that replacing S_{t-} with S_t cannot alter the Lebesgue integral over t, since $S_{t-}(\omega)$ differs from $S_t(\omega)$ for only countably many t. We obtain 7.19 by evaluating the last expectation using the distribution π_t of S_t and recalling that $\lambda\{0\} = 0$. □

At the time S crosses x, its left-limit G_x belongs to $[0, x]$ and its right-hand value D_x belongs to $[x, \infty)$. Thus, if the crossing is by a jump, the jump is either from somewhere in $[0, x]$ into (x, ∞) or from somewhere in $[0, x)$ to the point x. The following shows that the last possibility is improbable.

7.21 Corollary. *For x in $(0, \infty)$,*

$$\mathbb{P}\{\ G_x < x = D_x\ \} = 0.$$

Proof. Fix x. This is obvious when S is compound Poisson, because $D_x > x$ then. Suppose that S is not compound Poisson, and recall that, then, the potential measure is diffuse. From the preceding theorem, taking $f(t, y, z) = 1_{[0,x)}(y)\, 1_{\{x\}}(z)$, we get

$$\mathbb{P}\{\ G_x < x = D_x\ \} = \int_{[0,x)} U(dy)\ \lambda\{\, x - y\,\}.$$

Since λ is σ-finite, it can have $x - y$ as an atom for at most countably many y; let A be the set of such y in $[0, x)$. We have $U(A) = 0$ since U is diffuse. So, the last integral is zero as claimed. □

7.22 Corollary. *For every x in $\mathbb{R}_+$,*

7.23 $$\mathbb{P}\{\ G_x = D_x\ \} = \mathbb{P}\{\ D_x = x\ \} = 1 - \int_{[0,x]} U(dy)\ \lambda\,(x - y, \infty).$$

Proof. For $x = 0$, this is by direct checking; see Remark 7.20. Suppose $x > 0$. It follows from the last corollary that, on the event $\{D_x = x\}$, we have $G_x = x$ almost surely; hence,

$$\mathbb{P}\{ D_x = x \} = \mathbb{P}\{ G_x = x = D_x \} = \mathbb{P}\{ G_x = D_x \}.$$

This proves the first equality. The second is obtained by computing $\mathbb{P}\{D_x > x\}$ from 7.19 by taking $f(t, y, z) = 1_{(x,\infty)}(z)$. □

Consider the preceding corollary in light of Theorem 7.12. If $b > 0$, the potential measure admits a density u, and comparing 7.13 and 7.23, we see that the probability of drifting across x is

7.24 $$\mathbb{P}\{ G_x = D_x \} = \mathbb{P}\{ D_x = x \} = b\, u(x), \qquad x \in \mathbb{R}_+.$$

If $b = 0$ and λ finite, that is, if S is compound Poisson, then $D_x > x$ for every x; hence,

7.25 $$\mathbb{P}\{ G_x = D_x \} = \mathbb{P}\{ D_x = x \} = 1 - \int_{[0,x]} U(dy)\, \lambda(x - y, \infty) = 0,$$

for x in $\mathbb{R}_+$; and we see that 7.11 is true for every x as a by-product. Indeed, as remarked in 7.10, it can be shown that 7.25 is true for every x as long as $b = 0$. Here is an example.

7.26 Example. *Stable processes.* Suppose that S is the increasing stable process of Example 7.6. Recall that $\lambda(dx) = dx\, a / x^{a+1}\, \Gamma(1-a)$, which yielded the potential measure $U(dx) = dx / x^{1-a}\, \Gamma(a)$. Then, for $0 \le y \le x < z$, we see from 7.19 that

$$\begin{aligned} \mathbb{P}\{ G_x \in dy,\ D_x \in dz \} &= U(dy)\ \lambda(dz - x) \\ &= dy\ dz\ \frac{a}{\Gamma(a)\,\Gamma(1-a) y^{1-a} (z-y)^{1+a}} \\ &= dy\ dz\ \frac{a \sin \pi a}{\pi y^{1-a} (z-y)^{1+a}}. \end{aligned}$$

Integrating over y in $[0, x]$ and z in (x, ∞), we get $\mathbb{P}\{ D_x > x \} = 1$, confirming 7.25 and 7.11 once more.

Drifting across

We concentrate here on the distribution of T_x in the event x is crossed by drifting. Define

7.27 $$\mu_x(A) = \mathbb{P}\{ T_x \in A,\ G_x = D_x \}, \qquad x \in \mathbb{R}_+,\ A \in \mathcal{B}_{\mathbb{R}_+}.$$

If $b = 0$ then this is zero. Suppose that $b > 0$. Then, S is strictly increasing, which implies that $x \mapsto T_x$ is continuous, which in turn implies that

$x \mapsto \mu_x(A)$ is Borel measurable for each A in $\mathcal{B}_{\mathbb{R}_+}$. Hence, $(x, A) \mapsto \mu_x(A)$ is a transition kernel; it is bounded, since $\mu_x(\mathbb{R}_+) = b\,u(x)$ in view of 7.24, and u is bounded by $1/b$. The following identifies it.

7.28 THEOREM. *Suppose that $b > 0$. Then, μ is a transition kernel from $\mathbb{R}_+$ into $\mathbb{R}_+$ which satisfies*

7.29 $$dx\ \mu_x(dt) \ = \ dt\ \pi_t(dx)\ b, \qquad x \in \mathbb{R}_+,\ t \in \mathbb{R}_+.$$

Proof. Let $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be Borel. With $b > 0$, the form 7.1 of S shows that $dS_t(\omega) = b\,dt$ if $S_{t-}(\omega) = S_t(\omega)$. Hence,

$$\begin{aligned}\int_{\mathbb{R}_+} dt\ b\,f(t, S_t) &= \int_{\mathbb{R}_+} f(t, S_t)\ 1_{\{\,S_{t-}=S_t\,\}} dS_t \\ &= \int_{\mathbb{R}_+} dx\, f(T_x, D_x)\ 1_{\{\,G_x=D_x\,\}} \\ &= \int_{\mathbb{R}_+} dx\, f(T_x, x)\, 1_{\{\,G_x=D_x\,\}},\end{aligned}$$

where we used the time change $t = T_x$, the definitions 7.17 of G and D, and the observation that $D_x = x$ on $\{\,G_x = D_x\,\}$. Next, we take expectations on both sides; using the definition 7.27, we get

$$b\int_{\mathbb{R}_+} dt \int_{\mathbb{R}_+} \pi_t(dx)\ f(t, x) \ = \ \int_{\mathbb{R}_+} dx \int_{\mathbb{R}_+} \mu_x(dt)\ f(t, x).$$

This proves 7.29 since f is an arbitrary positive Borel function. □

7.30 REMARK. Fix $t > 0$. let $\nu(A)$ be the expected amount of time that S spends in the set A during the time interval $[0, t]$. Then, ν is a measure on $\mathbb{R}_+$ whose total mass is t. According to 7.29, ν is absolutely continuous with respect to the Lebesgue measure, and

$$\mu_x\,[\,0, t\,] \ = \ b\ \frac{\nu(dx)}{dx}, \qquad x \in \mathbb{R}_+.$$

7.31 EXAMPLE. Suppose that $S_t = bt + S^o_t$, where S^o is a gamma process with shape rate a and scale parameter c. Then,

$$\pi_t(dx) \ = \ dx\ \frac{e^{-c(x-bt)}\ c^{at}\ (x - bt)^{at-1}}{\Gamma(at)}, \qquad x > bt;$$

and

$$\mu_x(dt) \ = \ dt\ \frac{b\,c^{at}\ e^{-c(x-bt)}\ (x - bt)^{at-1}}{\Gamma(at)}\ 1_{(0,x)}(bt).$$

Exercises

7.32 *Compound Poisson.* Suppose that S is a compound Poisson process with an exponential jump size distribution, that is, its Lévy measure is $\lambda(dx) = ca\, e^{-ax}\, dx$ for some constants a and c in $(0, \infty)$. Show that the corresponding potential measure is

$$U(dx) \;=\; \frac{1}{c}\,\delta_0(dx) + \frac{a}{c}\,dx, \qquad\qquad x \in \mathbb{R}_+.$$

7.33 *Atoms of* π_t. Theorem 7.28 might suggest that, when $b > 0$, the distribution π_t is absolutely continuous. This is false: Suppose that $S_t = bt + N_t$ where N is Poisson with rate c. For fixed $x > 0$, then

$$\pi_t\{\,x\,\} \;=\; \mathbb{P}\{\,S_t = x\,\} \;=\; \mathbb{P}\{\,N_t \;=\; x - bt\,\},$$

which is strictly positive if $x - bt = n$ for some integer $n \geq 0$. In the positive direction, it is known that π_t is diffuse whenever the Lévy measure is infinite.

7.34 *Poisson with drift.* Suppose that $S_t \;=\; t + N_t$ where N is a Poisson process with rate 1. Fix $x > 0$. Show that

$$\{\,D_x = x\,\} \;=\; \bigcup_k \{\,T_x = x - k\,,\; N_{x-k} = k\,\} \;=\; \bigcup_k \{\,N_{x-k} = k\,\}$$

where the sum is over all integers k in $[0, x)$. Show that

$$u(x) \;=\; \mathbb{P}\{\,G_x = D_x = x\,\} \;=\; \sum_{k<x} \frac{e^{-(x-k)}\,(x-k)^k}{k!}.$$

Compute

$$\begin{aligned}\mu_x[0,t] &= \mathbb{P}\{\,T_x \leq t,\; G_x = D_x = x\,\}\\ &= \mathbb{P}\{\,T_x \leq t\,\} \;-\; \mathbb{P}\{\,T_x \leq t,\; G_x \neq D_x\,\}.\end{aligned}$$

7.35 *Stable process with index* $a = 1/2$. Suppose that S is stable with index $1/2$; then, $b = 0$ and the Lévy measure is $\lambda(dx) \;=\; dx\,(c/x^{a+1})\,1_{(0,\infty)}(x)$ for some constant c. Show that the distribution of $(\,G_x, D_x\,)$ is free of c. Use Example 7.26 to show that

$$\mathbb{P}\{\,G_x \in dy,\; D_x \in dz\,\} \;=\; dy\,dz\;\frac{1}{2\pi\,\sqrt{y\,(z-y)^3}}, \qquad y < x < z.$$

Show that, for $y < x < z$ again,

$$\mathbb{P}\{\,G_x < y,\; D_x > z\,\} \;=\; \frac{2}{\pi}\;\arcsin\sqrt{\frac{y}{z}}.$$

In particular, then, for $y < x$,

$$\mathbb{P}\{\,G_x < y\,\} \;=\; \mathbb{P}\{\,D_y > x\,\} \;=\; \frac{2}{\pi}\;\arcsin\sqrt{\frac{y}{x}}.$$

The distribution involved here is called the arcsine distribution; it is the beta distribution with index pair $(\frac{1}{2}, \frac{1}{2})$.

7.36 *Drifting.* In general, if $b > 0$, show that

$$\mathbb{P}\{ T_x > t,\ G_x = D_x \} \ = \ \pi_t\,[0,x] \ - \ \int_t^\infty du \int_{[0,x]} \pi_u(dy)\ \lambda\,[\,x - y, \infty\,).$$

7.37 *Laplace transforms.* Let $\psi(p) \ = \ bp + \int \lambda(dx)\,(1 - e^{-px})$, the Laplace exponent for S, for $p \geq 0$. Show that, for $p > 0$,

$$\int_{\mathbb{R}_+} dx\ e^{-px}\ \mathbb{P}\{ T_x > t \} = \frac{1}{p}\ e^{-t\psi(p)},$$

$$\int_{\mathbb{R}_+} dx\ e^{-px}\ \mathbb{P}\{ T_x > t,\, G_x = D_x \} = \frac{b}{\psi(p)}\ e^{-t\psi(p)},$$

$$\int_{\mathbb{R}_+} dx\ p\,e^{-px}\ \mathbb{E}\ T_x = \frac{1}{\psi(p)} \ = \ \hat{u}(p).$$

7.38 *Time changes.* Let $c\ :\ \mathbb{R}_+ \mapsto \mathbb{R}_+$ be a strictly increasing continuous function with $c(0) = 0$ and $\lim_{t\to\infty} c(t) = +\infty$. Define

$$\hat{S}_t \ = \ S_{c(t)}\,, \qquad t \in \mathbb{R}_+.$$

Then, $\hat{S}$ is a process with independent increments, but the stationarity of increments is lost unless $c(t) = c_0\,t$. Define $\hat{T}$, $\hat{G}$, $\hat{D}$ from the process $\hat{S}$ in the same manner that T, G, D are defined from S.

a) Show that $\hat{G}_x = G_x$ and $\hat{D}_x = D_x$ for all x.

b) Show that $c\,(\hat{T}_x) = T_x$; thus, $\hat{T}_x \ = \ a\,(T_x)$ where a is the functional inverse of c.

7.39 *Continuation.* Observe that the preceding results remain true when c is replaced by a stochastic clock C whose paths $t \mapsto C(\omega, t)$ satisfy the conditions on c for every ω.

Chapter VIII

Brownian Motion

This chapter is on Brownian motions on the real line $\mathbb{R}$ with a few asides on those in $\mathbb{R}^d$. We concentrate on the Wiener process, the standard Brownian motion.

Section 1 introduces Brownian motions, indicates their connections to martingales, Lévy processes, and Gaussian processes, and gives several examples of Markov processes closely related to Brownian motions. Section 2 is on the distributions of hitting times and on the arcsine law for the probability of avoiding the origin. Section 3 treats the hitting times as a process; the process turns out to be an increasing pure-jump Lévy process that is stable with index $1/2$.

The Wiener process W and its running maximum M are studied jointly in Section 4; it is shown that $M - W$ is a reflected Brownian motion and that $2M - W$ is a Bessel process. The relationship of M to $M - W$ is used to introduce the local time process for W; this is put in Section 5 along with the features of the zero-set for W. Brownian excursions are taken up in Section 6; the Poisson random measure of excursions is described, and the major arcsine law (on time spent on the positive half-line) is derived as an application.

Section 7 is on the fine properties of Brownian paths: total variation, quadratic variation, Hölder continuity, and the law of the iterated logarithm. Finally, in Section 8, we close the circle by showing that Brownian motions do exist; we give two constructions, one due to Lévy and one using Kolmogorov's theorem on continuous modifications.

1 Introduction

The aim is to introduce Brownian motions and Wiener processes. We start with an elementary definition and enhance it to its modern version. We shall also consolidate some results from the chapters on martingales

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_8,

and Lévy processes. Finally we describe several Markov processes which are closely related to Brownian motions. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is the probability space in the background.

1.1 DEFINITION. *A stochastic process* $X = (X_t)_{t\in\mathbb{R}_+}$ *with state space* $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ *is called a Brownian motion if it is continuous and has stationary independent increments. A process* $W = (W_t)_{t\in\mathbb{R}_+}$ *is called a Wiener process if it is a Brownian motion with*

1.2 $$W_0 = 0, \qquad \mathbb{E}\, W_t = 0, \qquad \mathrm{Var}\ W_t = t, \quad t \in \mathbb{R}_+.$$

Let X be a Brownian motion. Then, $(X_t - X_0)_{t\in\mathbb{R}_+}$ is a continuous Lévy process. It follows from the characterization of such processes (see Theorem VII.4.2) that X has the form

1.3 $$X_t = X_0 + at + bW_t, \qquad t \in \mathbb{R}_+,$$

where a and b are constants in $\mathbb{R}$ and $W = (W_t)$ is a Wiener process independent of X_0. The constant a is called the *drift* rate, and b the *volatility* coefficient. The case $b = 0$ is degenerate and is excluded from further consideration.

Gaussian connection

Let W be a Wiener process. Its every increment $W_{s+t} - W_s$ has the Gaussian distribution with mean 0 and variance t:

1.4 $$\mathbb{P}\{W_{s+t} - W_s \in B\} = \mathbb{P}\{W_t \in B\} = \int_B dx \frac{e^{-x^2/2t}}{\sqrt{2\pi t}}, \quad t > 0;$$

see Theorem VII.4.2 *et seq*. This implies, via the independence of the increments over disjoint intervals, that the random vector $(W_{t_1}, \ldots, W_{t_n})$ has the n-dimensional Gaussian distribution with

$$\mathbb{E}\, W_{t_i} = 0, \qquad \mathrm{Cov}(W_{t_i}, W_{t_j}) = t_i, \qquad 1 \le i \le j \le n,$$

for arbitrary integers $n \ge 1$ and times $0 \le t_1 < \cdots < t_n$. Conversely, if $(W_{t_1}, \ldots, W_{t_n})$ has the n-dimensional Gaussian distribution described, then the increments $W_{t_1},\ W_{t_2} - W_{t_1},\ \ldots,\ W_{t_n} - W_{t_{n-1}}$ are independent Gaussian variables with mean 0 and respective variances $t_1,\ t_2 - t_1,\ \ldots,\ t_n - t_{n-1}$. These remarks prove the following.

1.5 THEOREM. *Let* $W = (W_t)$ *be a process with state space* $\mathbb{R}$. *It is a Wiener process if and only if it is continuous and is a Gaussian process with mean 0 and*

$$\mathrm{Cov}\ (W_s, W_t) = s \wedge t, \qquad s, t \in \mathbb{R}_+.$$

The preceding theorem is often useful in showing that a given process is Wiener; see the next theorem for an instance of its use. It also raises an interesting question: Does Brownian motion exist? After all, the probability law of a Gaussian process is determined completely by its mean and covariance functions; how do we know that we can satisfy the further condition that its paths be continuous? We shall give two proofs of its existence in Section 8.

Symmetry, scaling, time inversion

1.6 THEOREM. *Let W be a Wiener process. Then, the following hold:*

a) Symmetry. The process $(-W_t)_{t\in\mathbb{R}_+}$ is again a Wiener process.

b) Scaling. $\hat{W} = (c^{-1/2}W_{ct})_{t\in\mathbb{R}_+}$ is a Wiener process for each fixed c in $(0,\infty)$, that is, W is stable with index 2.

c) Time inversion. Putting $\tilde{W}_0 = 0$ and $\tilde{W}_t = tW_{1/t}$ for $t > 0$ yields a Wiener process $\tilde{W} = (\tilde{W}_t)_{t\in\mathbb{R}_+}$.

Proof. Symmetry and scaling properties are immediate from Definition 1.1 for Wiener processes. To show (c), we start by noting that $\{\tilde{W}_t : t > 0\}$ is a continuous Gaussian process with mean 0 and $\text{Cov}(\tilde{W}_s,\ \tilde{W}_t) = s \wedge t$ for $s, t > 0$. Thus, the claim (c) will follow from Theorem 1.5 once we show that $\tilde{W}$ is continuous at time 0, that is, almost surely,

1.7 $$\lim_{t\downarrow 0} t\ W_{1/t} = 0.$$

Equivalently, we shall show that $W_t/t \to 0$ almost surely as $t \to \infty$. To this end, we start by noting that, if $n \geq 0$ is an integer and $n < t \leq n+1$,

1.8 $$\left|\frac{1}{t}\ W_t\right| \leq \frac{1}{n}\left|W_n + (W_t - W_n)\right| \leq \left|\frac{1}{n}W_n\right| + \frac{1}{n}\sup_{0\leq s\leq 1}\left|W_{n+s} - W_n\right|.$$

By the strong law of large numbers, $W_n/n \to 0$ almost surely, since W_n is the sum of n independent copies of W_1, and $\mathbb{E}\,W_1 = 0$. On the other hand, by Kolmogorov's inequality in continuous time (Lemma VII.1.39),

$$\mathbb{P}\left\{\frac{1}{n}\sup_{0\leq s\leq 1}\left|W_{n+s} - W_n\right| > \varepsilon\right\} \leq \frac{1}{n^2\varepsilon^2}\mathbb{E}\left|W_{n+1} - W_n\right|^2 = \frac{1}{n^2\varepsilon^2}$$

for each fixed $\varepsilon > 0$. Since $\Sigma\ 1/n^2$ is finite, Borel–Cantelli lemma (III.2.6) applies to show that, as $n \to \infty$, the very last term in 1.8 goes to 0 almost surely. Hence, $W_t/t \to 0$ almost surely as $t \to \infty$, and the proof is complete. □

In connection with the stability property 1.6b, we recall from Exercise VII.2.36 the following converse: if a continuous Lévy process is stable with index 2, then it necessarily has the form cW for some fixed constant c and

some Wiener process W. As to the property 1.6c, time inversion, we remark at least two of its uses: first, the oscillatory behavior of a Wiener process near the time origin can be translated to its behavior for large times; second, conditioning on future values can be translated to become conditioning on the past. The following illustrates the latter point.

1.9 EXAMPLE. Let W be a Wiener process. For $0 < s < t$, consider the conditional distribution of W_s given that $W_t = x$. Instead of the direct approach, it is easier to use the time inversion property: The conditional distribution sought is that of $sW_{1/s}$ given that $tW_{1/t} = x$, which is the same as the distribution of $s(W_{1/s} - W_{1/t}) + \frac{sx}{t}$, which is Gaussian with mean sx/t and variance $s^2(1/s - 1/t) = s(1 - s/t)$. See Exercise 1.29 also.

Martingale connection

Let $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration over $(\Omega, \mathcal{H}, \mathbb{P})$, and let $W = (W_t)_{t\in\mathbb{R}_+}$ be a continuous process, adapted to $\mathcal{F}$, and having $W_0 = 0$. Recall Definition V.2.15: the process W is Wiener with respect to $\mathcal{F}$ if, for every t and u in $\mathbb{R}_+$, the increment $W_{t+u} - W_t$ is independent of $\mathcal{F}_t$ and has the Gaussian distribution with mean 0 and variance u.

If W is Wiener with respect to $\mathcal{F}$, then it is such in the sense of Definition 1.1 as well. Conversely, if W is Wiener in the sense of 1.1, then it is Wiener with respect to the filtration $\mathcal{G}^o$ generated by itself, and also with respect to the filtration $\mathcal{G}$, the augmentation of $\mathcal{G}^o$.

The following collects together characterizations in Proposition V.2.17 and Theorem V.2.19; see also Proposition V.6.21, Lemma V.6.22, and all the proofs. Recall that W is continuous, has $W_0 = 0$, and is adapted to the filtration $\mathcal{F}$.

1.10 THEOREM. *The following are equivalent:*

a) W is a Wiener process with respect to $\mathcal{F}$.

b) For each r in $\mathbb{R}$, the process $\{\exp(rW_t - \frac{1}{2}r^2t) : t \in \mathbb{R}_+\}$ is an $\mathcal{F}$-martingale.

c) The processes W and $\{W_t^2 - t : t \in \mathbb{R}_+\}$ are $\mathcal{F}$-martingales.

d) For every twice-differentiable function $f : \mathbb{R} \mapsto \mathbb{R}$ that is bounded along with its first derivative f' and second derivative f'', the process

$$M_t = f \circ W_t - \frac{1}{2}\int_0^t ds\ f'' \circ W_s, \qquad t < \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale.

The preceding theorem is on the characterization of Wiener processes as martingales. Indeed, the connections between them run deep in both directions. In particular, it is known that every continuous martingale is obtained from a Wiener process by a random time change.

Wiener on a stochastic base

This is to re-introduce Wiener processes in the modern setup for Lévy processes; this is a repetition of Definitions VII.3.1 and VII.3.3 *et seq.* for this particular case.

Recall that a *stochastic base* is a collection $(\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$, where $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space, $\mathcal{F} = (\mathcal{F}_t)$ is an augmented right-continuous filtration, and $\theta = (\theta_t)$ is a semigroup of shift operators on Ω (each θ_t maps Ω into Ω, we have $\theta_0\omega = \omega$ for all ω, and $\theta_u \circ \theta_t = \theta_{t+u}$ for all t and u in $\mathbb{R}_+$).

1.11 Definition. *A process $W = (W_t)$ is said to be Wiener on a stochastic base $(\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$ if it is a Wiener process with respect to $\mathcal{F}$ and is additive with respect to θ, the latter meaning that*

$$W_{t+u} = W_t + W_u \circ \theta_t, \qquad t, u \in \mathbb{R}_+.$$

The shift operators and additivity are useful for turning heuristic feelings into rigorous statements; for instance, $W_u \circ \theta_t$ is the increment over the future interval of length u when the present time is t, and the future is totally independent of the past. The right-continuity of $\mathcal{F}$ is essential for certain times to be $\mathcal{F}$-stopping times; augmentedness is for technical comfort. There is no loss of generality in all this: Every Wiener process in the sense of Definition 1.1 is equivalent to one in the sense of the preceding definition.

Brownian motions X on a stochastic base are defined similarly, except for the way the shifts work:

1.12 $$X_u \circ \theta_t = X_{t+u}, \qquad t, u \in \mathbb{R}_+.$$

This is equivalent to the additivity of W in the characterization 1.3 for X. See Figure 9 on page 341 for additivity.

Strong Markov property

Let $(\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$ be a stochastic base, and W a Wiener process over it. Let $\mathcal{G}^o = (\mathcal{G}_t^o)$ be the filtration generated by W, and $\mathcal{G}$ the augmentation of $\mathcal{G}^o$. Recall from Theorem VII.3.20 that $\mathcal{G}$ is right-continuous in addition to being augmented; it can replace $\mathcal{F}$ if needed. In particular, Blumenthal's zero-one law holds: every event in $\mathcal{G}_0$ has probability zero or one.

The following is the strong Markov property, Theorem VII.3.10, for the special Lévy process W, we re-state it here for reasons of convenience. As usual, we write $\mathbb{E}_T$ for $\mathbb{E}(\cdot|\mathcal{F}_T)$.

1.13 Theorem. *Let T be an $\mathcal{F}$-stopping time. Then, for every bounded variable V in $\mathcal{G}_\infty$,*

$$\mathbb{E}_T(V \circ \theta_T)1_{\{T<\infty\}} = (\mathbb{E}V)1_{\{T<\infty\}}.$$

In particular, if $T < \infty$, the process $W \circ \theta_T = (W_{T+u} - W_T)_{u \in \mathbb{R}_+}$ is independent of $\mathcal{F}_T$ and is again a Wiener process.

Let U be a random time determined by the past $\mathcal{F}_T$ and consider $W_{T+U} - W_T$. Since $W \circ \theta_T$ is independent of $\mathcal{F}_T$, we may treat U as if it is fixed. We list the result next and give a direct proof. The heuristic idea is simpler, but requires some sophistication in its execution; see Exercise 1.31.

1.14 THEOREM. *Let T be an $\mathcal{F}$-stopping time, and let U be a positive real-valued variable belonging to $\mathcal{F}_T$. Let f be a bounded Borel function on $\mathbb{R}$, and define $g(u) = \mathbb{E} f \circ W_u,\ u \in \mathbb{R}_+$. Then,*

1.15 $$\mathbb{E}_T\ f\,(W_{T+U} - W_T)\ 1_{\{T<\infty\}} = g(U)\ 1_{\{T<\infty\}}.$$

Proof. a) The collection of f for which 1.15 holds is a monotone class. Thus, it is enough to show 1.15 for f that are bounded continuous. Fix f such, and note that the corresponding g is bounded and continuous in view of the continuity of W and the bounded convergence theorem for expectations.

b) Suppose that U is simple, say, with values in a finite subset D of $\mathbb{R}_+$. Since U is $\mathcal{F}_T$-measurable, $\{U = u\}$ is in $\mathcal{F}_T$ for each u in D. Thus,

$$\begin{aligned} &\mathbb{E}_T f\,(W_{T+U} - W_T)\, 1_{\{U=u, T<\infty\}} \\ &\quad = \mathbb{E}_T\ 1_{\{U=u,T<\infty\}}\ f(W_{T+u} - W_T) = g(u) 1_{\{U=u\}} 1_{\{T<\infty\}}, \end{aligned}$$

where we used the strong Markov property 1.13 at the last step. Summing both sides over all u in D yields 1.15. So, 1.15 holds for simple U.

c) In general, U is the limit of an increasing sequence (U_n) of simple variables in $\mathcal{F}_T$. Write 1.15 for U_n and take limits on both sides as $n \to \infty$. On the right side, the continuity of g shows that the limit is the right side of 1.15. On the left side, the continuity of W and f, together with the boundedness of f, imply that the limit is the left side of 1.15. □

Wiener and Brownian motion in $\mathbb{R}^d$

Let $W = (W_t)$ be a process with state space $\mathbb{R}^d$. It is called a *d-dimensional Wiener* process, or a *Wiener process in* $\mathbb{R}^d$, if its components $W^{(1)}, \ldots, W^{(d)}$ are independent Wiener processes. Then, W is a continuous Lévy process in $\mathbb{R}^d$ whose every increment $W_{s+t} - W_s$ has the d-dimensional Gaussian distribution with mean 0 and covariance matrix tI, the matrix I being the identity matrix in d-dimensions. So, for Borel subsets B of $\mathbb{R}^d$,

1.16 $$\mathbb{P}\,\{W_{s+t} - W_s \in B\} = \int_B dx\ \frac{e^{-|x|^2/2t}}{(2\pi t)^{d/2}},$$

where $|x|$ is the length of the vector x in $\mathbb{R}^d$, and the integral is with respect to the Lebesgue measure on $\mathbb{R}^d$.

The properties of symmetry, 2-stability, and time inversion remain true for the d-dimensional case. Moreover, symmetry is extended to *isotropy*, invariance of the law of W under rotations and reflections: the probability laws of W and gW are the same for every orthogonal matrix g.

Brownian motions in $\mathbb{R}^d$ are defined as the ones in $\mathbb{R}$, except that the state space is $\mathbb{R}^d$ now. Every Brownian motion X in $\mathbb{R}^d$ is related to a Wiener process in $\mathbb{R}^d$ by the formula 1.3, but here a is a fixed vector in $\mathbb{R}^d$ and b is a fixed $d \times d$ matrix.

Markov processes

Brownian motions are the fundamental objects from which all continuous Markov processes are constructed. Several examples occur naturally as parts of the theory of Brownian motions. It will be convenient to provide a working definition for our current purposes and give several examples; see the next chapter for more.

Over some probability space $(\Omega, \mathcal{H}, \mathbb{P})$, let $X = (X_t)_{t\in\mathbb{R}_+}$ be a stochastic process with some state space $(E, \mathcal{E})$ and suppose that it is adapted to some filtration $\mathcal{F}^o = (\mathcal{F}^o_t)$. For each t, let P_t be a markovian kernel on $(E, \mathcal{E})$, that is, a transition kernel from $(E, \mathcal{E})$ into $(E, \mathcal{E})$ with $P_t(x, E) = 1$ for every x in E. Then, X is said to be an $\mathcal{F}^o$-*Markov process* with *transition semigroup* $(P_t)_{t\in\mathbb{R}_+}$ if

1.17 $$\mathbb{P}\left\{X_{s+t} \in B \,|\, \mathcal{F}^o_s\right\} = P_t\left(X_s, B\right), \qquad s, t \in \mathbb{R}_t,\ B \in \mathcal{E}.$$

The term "Markov process" without the mention of a filtration refers to the case where $\mathcal{F}^o$ is the filtration generated by the process itself.

The condition 1.17 implies that the Markovian kernels $P_t,\ t \in \mathbb{R}_+$, do indeed form a *semigroup*: $P_s\ P_t = P_{s+t}$ for s, t in $\mathbb{R}_+$, or, more explicitly,

1.18 $$P_{s+t}(x, B) = \int_E P_s(x, dy) P_t(y, B), \qquad s, t \in \mathbb{R}_+,\ x \in E,\ B \in \mathcal{E}.$$

Imagine a particle whose motion in E is represented by the process X. The defining property 1.17 means, in particular, that

$$P_t(x, B) = \mathbb{P}\{X_{s+t} \in B \,|\, X_s = x\}, \qquad x \in E,\ B \in \mathcal{E}.$$

The independence of this conditional probability from the time parameter s is referred to as *time-homogeneity* for X. Repeated use of 1.17 implies that, given the past $\mathcal{F}^o_s$, the conditional law of the future motion $\{X_{s+t} : t \in \mathbb{R}_+\}$ depends only on the present state X_s. A similar reasoning shows that the probability law of the process X is determined by its transition semigroup and its initial distribution (the distribution of X_0).

Examples

1.19 *Brownian motion in* $\mathbb{R}^d$. Let $X_t = X_0 + W_t,\ t \in \mathbb{R}_+$, where W is a Wiener process in $\mathbb{R}^d$ independent of X_0. Then, X is a Markov process with state space $\mathbb{R}^d$. Its transition semigroup is given as (see 1.16)

1.20 $$P_t(x, dy) = dy\frac{e^{-|y-x|^2/2t}}{(2\pi t)^{d/2}}, \qquad x, y \in \mathbb{R}^d.$$

In particular, W is a Markov process (with initial state $W_0 = 0$) with the same transition semigroup.

1.21 *Reflected Brownian motion.* Let $X = X_0 + W$ be a standard Brownian motion in $\mathbb{R}$, with initial state X_0. Define $R = |X|$, that is, R_t is the absolute value of X_t. Then, R is a Markov process with state space $\mathbb{R}_+$. To compute its transition semigroup (P_t), we start by noting that (see 1.20 with $d = 1$)

$$\mathbb{P}\{R_{s+t} \in dy | X_s = x\} = dy\left[\frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} + \frac{e^{-(-y-x)^2/2t}}{\sqrt{2\pi t}}\right]$$

for x in $\mathbb{R}$ and y in $\mathbb{R}_+$. The right side remains the same whether x is positive or negative. Thus,

$$P_t(x, dy) = dy\left[\frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} + \frac{e^{-(y+x)^2/2t}}{\sqrt{2\pi t}}\right], \quad x,\ y \in \mathbb{R}_+.$$

1.22 *Bessel processes of index d.* This is the generalization of the preceding to higher dimensional Brownian motions. Let W be a Wiener process in $\mathbb{R}^d$ and define $R = |W|$, that is,

$$R_t = \sqrt{\left(W_t^{(1)}\right)^2 + \cdots + \left(W_t^{(d)}\right)^2}, \qquad t + \mathbb{R}_+.$$

Then, we call R a *Bessel process of index d*; some authors call it a *Bessel process of order* $\nu = d/2 - 1$, or *radial Brownian motion* in $\mathbb{R}^d$. It is a Markov process with state space $\mathbb{R}_+$; we shall show this. The case $d = 3$ plays an interesting role in describing the excursions of the one-dimensional Wiener away from the origin; we shall compute its transition semigroup explicitly.

For arbitrary dimension d, fixed, let B denote the closed unit ball in $\mathbb{R}^d$ and S its boundary, the unit sphere. For r in $\mathbb{R}_+$, then, $Br = \{xr :\ x \in B\}$ is the closed ball of radius r centered at the origin. From 1.16, we get

$$\mathbb{P}\{W_{s+t} \in Br \,|\, W_s = x\} = \int_{Br} dy\ \frac{e^{-|y-x|^2/2t}}{(2\pi t)^{d/2}}.$$

The left side remains unchanged if x, B, W are replaced with gx, gB, gW respectively, where g is some orthogonal transformation. But $gB = B$ since B is a ball centered at the origin, and gW has the same law as W by isotropy.

Hence, if $|x| = q$, choosing g such that $gx = (q,\ 0, \ldots, 0)$, we see that the left side is a function of $|x| = q$ only. Since $|W_s| = R_s$ and $\{W_{s+t} \in Br\} = \{R_{s+t} \le r\}$, we have shown that

$$1.23 \qquad \mathbb{P}\{R_{s+t} \le r \,|R_s = q\} = \int_{Br} dy\ \frac{e^{-|y-x|^2/2t}}{(2\pi t)^{d/2}}, \qquad q, r \in \mathbb{R}_+,$$

with $x = (q, 0, \ldots, 0)$ on the right side. Moreover, R_{s+t} is conditionally independent of $(W_u)_{u \le s}$ given W_s, and $(W_u)_{u \le s}$ determines $(R_u)_{u \le s}$. Thus, R_{s+t} is conditionally independent of $(R_u)_{u \le s}$ given W_s, and we have just seen that the conditional distribution of R_{s+t} given W_s is determined by $|W_s| = R_s$. Hence, R is Markov.

To evaluate the integral on the right side of 1.23, we turn to spherical coordinates. Write $y = ru$ with $u = (u_1, \ldots, u_d)$ on the unit sphere S. For $x = (q, 0, \ldots, 0)$, then, $|y - x|^2 = q^2 + r^2 - 2qr\ u_1$. Hence,

$$1.24 \qquad P_t(q, dr) = dr \quad r^{d-1}\ \frac{e^{-\left(q^2+r^2\right)/2t}}{(2\pi t)^{d/2}} \int_S \sigma(du)\, e^{qru_1/t},$$

where σ is the surface measure on S. The integral over S can be expressed in terms of modified Bessel functions (see Exercises 1.33 and 1.34), and hence the term Bessel process for R.

The surface integral is easy to evaluate when $d = 3$. We recall a result from elementary geometry: For spherical zones between two parallel planes that cut through S, the area is proportional to the distance h between the planes. So,

$$\int_S \sigma(du)\, e^{pu_1} = 2\pi \int_{-1}^{1} dh\ e^{ph} = \frac{2\pi}{p}\left(e^p - e^{-p}\right)$$

for $p > 0$, and the integral is the surface area 4π for $p = 0$. Putting this into 1.24 with $p = qr/t$, we see that, when $d = 3$,

$$1.25 \quad P_t(q, dr) = dr \frac{r}{q}\left[\frac{e^{-(r-q)^2/2t}}{\sqrt{2\pi t}} - \frac{e^{-(r+q)^2/2t}}{\sqrt{2\pi t}}\right] \qquad \textit{if } q > 0,\ r \ge 0,$$

and

$$1.26 \qquad P_t(q, dr) = dr \cdot \frac{2r^2 e^{-r^2/2t}}{\sqrt{2\pi t^3}} \qquad \textit{if } q = 0,\ r \ge 0,$$

We shall see later that, for almost every ω, we have $R_t(\omega) > 0$ for all $t > 0$; see 4.17 and thereabouts.

Exercises and complements

1.27 *Time reversal.* Let W be a Wiener process (on $\mathbb{R}$). Show that the probability laws of $\{W_t : 0 \le t \le 1\}$ and $\{W_1 - W_{1-t} :\ 0 \le t \le 1\}$ are the same. Hint: They are both Gaussian processes.

1.28 *Brownian bridge.* Let W be a Wiener process and define

$$X_t = W_t - tW_1, \qquad 0 \le t \le 1.$$

Observe that $X_0 = X_1 = 0$ and hence the name for the process $X = \{X_t : 0 \le t \le 1\}$. Obviously, X is a continuous Gaussian process. Compute its covariance function.

1.29 *Continuation.* Show that the probability law of X is the same as the conditional law of $\{W_t : 0 \le t \le 1\}$ given that $W_1 = 0$. In other words, show that, for $0 < t_1 < \ldots < t_n < 1$,

$$\mathbb{P}\left\{W_{t_1} \in dx_1, \ldots, W_{t_n} \in dx_n | W_1 = 0\right\} = \mathbb{P}\left\{X_{t_1} \in dx_1, \cdots, X_{t_n} \in dx_n\right\}.$$

Hint: Use time inversion (see Example 1.9) to show that the left side is an n-dimensional Gaussian distribution just as the right side, and compare their covariance matrices.

1.30 *Wiener space.* This is a special case of Exercise VII.3.24. Let W be a Wiener process on some probability space $(\Omega, \mathcal{H}, \mathbb{P})$. Let $C = C(\mathbb{R}_+ \mapsto \mathbb{R})$, the space of continuous functions from $\mathbb{R}_+$ into $\mathbb{R}$. On it, we put the topology of uniform convergence on compacts: a sequence (w_n) in C converges to w in C in this topology if $\sup_{s \le t} |w_n(s) - w(s)| \to 0$ as $n \to \infty$ for every $t < \infty$. It can be shown that the Borel σ-algebra $\mathcal{B}_C$ corresponding to this topology is the same as the σ-algebra generated by the coordinate process $\{X_t : t \in \mathbb{R}_+\}$, where $X_t(w) = w(t)$ for every w in C. Let $\mathcal{G}^0_\infty$ be the σ-algebra generated by $\{W_t : t \in \mathbb{R}_+\}$.

For each ω in Ω, the path $W(\omega) : t \mapsto W_t(\omega)$ is a point in C. Show that the mapping $\omega \mapsto W(\omega)$ is measurable with respect to $\mathcal{G}^o_\infty$ and $\mathcal{B}_C$.

Let $\mathbb{Q} = \mathbb{P} \circ W^{-1}$, the distribution of W, where W is regarded as a random variable taking values in $(C, \mathcal{B}_C)$. Then, $\mathbb{Q}$ is the probability law of the Wiener process W. The probability space $(C, \mathcal{B}_C, \mathbb{Q})$ is called the Wiener space, and $\mathbb{Q}$ the Wiener measure. Finally, X is a Wiener process on $(C, \mathcal{B}_C, \mathbb{Q})$ and is called the *canonical Wiener process.*

1.31 *Alternative proof for Theorem 1.14.* Assume that $T < \infty$. Define $Y_t = W_t \circ \theta_T = W_{T+t} - W_T$. By the strong Markov property, the process $Y = (Y_t)$ is independent of $\mathcal{F}_T$ and is a Wiener process. Regard Y as a random variable taking values in $(C, \mathcal{B}_C)$, and consider $Y_U = W_{T+U} - W_T$. Since U is in $\mathcal{F}_T$ and Y is independent of $\mathcal{F}_T$, Exercise IV.2.27 is applicable. Conclude that 1.15 holds since

$$g(u) = \mathbb{E}\ f(W_u) = \mathbb{E}\ f(Y_u).$$

1.32 *Geometric Brownian motion.* Let W be a Wiener process and put

$$X_t = X_0 \exp(at + bW_t), \qquad t \in \mathbb{R}_+,$$

for fixed constants a and b in $\mathbb{R}$. Show that X is a Markov process.

1.33 *Bessel process of index $d = 2$.* Let R be as in Example 1.22 but with $d = 2$. It is a Markov process with state space $\mathbb{R}_+$. To compute its semigroup (P_t), we use 1.24 with $d = 2$, in which case S becomes the unit circle in $\mathbb{R}^2$. Since

$$\int_S \sigma(du) e^{pu_1} = \int_0^{2\pi} da\ e^{p\cos a} = 2\pi \sum_{k=0}^{\infty} \frac{(p/2)^{2k}}{(k!)^2} = 2\pi\ I_0(p),$$

one obtains

$$P_t(q, dr) = dr\ \frac{r}{t}\ e^{-(q^2+r^2)/2t}\ I_0\left(\frac{qr}{t}\right), \qquad q, r \leq 0.$$

Here, I_0 is called the modified Bessel function of order 0, and hence the alternative name "Bessel process of order 0" for this R.

1.34 *Bessel processes.* Let R be as in Example 1.22 with arbitrary index $d \geq 2$. For $q > 0$ and $r \geq 0$, the formula 1.24 yields

$$P_t(q, dr) = dr \cdot \frac{q}{t} \left(\frac{r}{q}\right)^{d/2} e^{-\left(q^2+r^2\right)/2t}\ I_{d/2-1}\left(\frac{qr}{t}\right),$$

where I_ν is the modified Bessel function of order ν:

$$I_\nu\ (p) = \sum_{k=0}^{\infty} \frac{(p/2)^{2k+\nu}}{k!\Gamma(k+\nu+1)}, \qquad p \geq 0.$$

1.35 *Ornstein-Uhlenbeck process.* Let W be a Wiener process and write $W(t)$ for W_t. Let a and b be strictly positive constants, and define

1.36 $$X_t = X_0\ e^{-at} + b\ e^{-at}\ W\left(e^{2at} - 1\right),\ t \in \mathbb{R}_+,$$

where X_0 is independent of W.

a) Show that X defined by 1.36 is a Markov process with state space $\mathbb{R}$. It is also a Gaussian process if $X_0 = x$ fixed, or if X_0 is Gaussian.

b) Show that, as $t \to \infty$, the distribution of X_t converges weakly to the Gaussian distribution with mean 0 and variance b^2. If X_0 is Gaussian with mean 0 and variance b^2, and X_0 is independent of W, then X_t has the same distribution as X_0 for all t.

2 Hitting Times and Recurrence Times

Let $(\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$ be a stochastic base, and W a Wiener process on it; see Definition 1.11. By redefining $t \mapsto W_t(\omega)$ for a negligible set of ω if necessary, we may and do assume that $W_0(\omega) = 0$ and $t \mapsto W_t(\omega)$ is continuous for

every ω in Ω. As before, we let $\mathcal{G}^o$ be the filtration generated by W, and $\mathcal{G}$ its augmentation. We are interested in the hitting times

$$T_a(\omega) = \inf\{t > 0 : \ W_t(\omega) > a\}, \qquad a \in \mathbb{R}_+, \omega \in \Omega. \tag{2.1}$$

It follows from general theorems that each T_a is a stopping time of $\mathcal{G}$, and its Laplace transform can be obtained by martingale techniques; see Chapter V for these. However, it is enjoyable to do the treatment once more and obtain the distribution directly by Markovian techniques.

Fix a in $\mathbb{R}_+$. For ω in Ω and $t > 0$, we have $T_a(\omega) < t$ if and only if $W_r(\omega) > a$ for some rational number r in $(0,t)$; this is because W is continuous and $W_0 = 0$. Since W_r is in $\mathcal{G}^0_t$ for each such r, it follows that the event $\{T_a < t\}$ belongs to $\mathcal{G}^0_t$. Hence, by Theorem V.7.4, T_a is a stopping time of the filtration $\left(\mathcal{G}^0_{t+}\right)$ and, therefore, of the finer filtrations $\mathcal{G} = (\mathcal{G}_t)$ and $\mathcal{F} = (\mathcal{F}_t)$.

Behavior at the origin

According to Blumenthal's zero-one law, every event in $\mathcal{G}_0$ has probability zero or one. The following is an application of it.

2.2 PROPOSITION. *Almost surely, $T_0 = 0$.*

Proof. The event $\{T_0 = 0\}$ belongs to $\mathcal{G}_0$ and, thus, has probability 0 or 1. To decide which, note that $\{W_t > 0\}$ has probability ½ and implies the event $\{T_0 < t\}$ for every $t > 0$. Thus, $\mathbb{P}\{T_0 < t\} \geq$ ½ for every $t > 0$, and letting $t \to 0$ concludes the proof. □

The preceding proposition is deeper than it appears. Considering the definition 2.1 for $a = 0$ carefully, we see that the following picture holds for almost every ω: For every $\varepsilon > 0$ there is $u < \varepsilon$ such that $W_u(\omega) > 0$; there is also $s < \varepsilon$ such that $W_s(\omega) < 0$, this being by symmetry (see 1.6a). Taking ε of the second phrase to be the time u of the preceding one, and recalling the continuity of the paths, we conclude that for every $\varepsilon > 0$ there are $0 < s < t < u < \varepsilon$ such that $W_s(\omega) < 0$, $W_t(\omega) = 0$, $W_u(\omega) > 0$. Iterating the argument with s replacing ε yields the following.

2.3 COROLLARY. *For almost every ω, there are times $u_1 > t_1 > s_1 > u_2 > t_2 > s_2 > \ldots$ with limit 0 such that, for each n,*

$$W_{u_n}(\omega) > 0, \quad W_{t_n}(\omega) = 0, \ W_{s_n}(\omega) < 0.$$

Thus, the Wiener path $W(\omega)$ is highly oscillatory. Starting with $W_0\ (\omega) = 0$, the path spends no time at 0; it crosses over and under 0 at least infinitely many times during the time interval $(0, \varepsilon)$, however small $\varepsilon > 0$ may be. This statement has an interesting counterpart for large times obtained by time inversion, by applying 2.3 to the Wiener process of 1.6c.

2.4 Corollary. *For almost every ω there exist times $u_1 < t_1 < s_1 < u_2 < t_2 < s_2 < \ldots$ with limit $+\infty$ such that*

$$\lim W_{s_n}(\omega) = -\infty, \quad \lim W_{u_n}(\omega) = +\infty,$$

and $W_{t_n}(\omega) = 0$ for every n; in particular, the set $\{t \in \mathbb{R}_+ : W_t(\omega) = 0\}$ is unbounded.

We shall see shortly that the Wiener particle touches every point a in $\mathbb{R}$, and its path oscillates in the vicinity of a just as it does in the vicinity of the point 0.

Distribution of T_a

We start with a useful formula based on the strong Markov property and, more particularly, on Theorem 1.14. For its statement, it will be convenient to introduce the *Gaussian kernel*

2.5 $$G(t, B) = \mathbb{P}\{W_t \in B\} = \int_B dx \, \frac{e^{-x^2/2t}}{\sqrt{2\pi t}}, \qquad t \in \mathbb{R}_+, \; B \in \mathcal{B}_{\mathbb{R}},$$

with $G(0, B)$ interpreted as $I(0, B)$ since $W_0 = 0$. Recall that $x + yB$ is the set of points $x + yz$ in $\mathbb{R}$ with z in B.

2.6 Lemma. *For t and a in $\mathbb{R}_+$, and B a Borel subset of $\mathbb{R}$,*

$$\mathbb{P}\{T_a \le t, W_t \in B\} = \mathbb{E}\, G(t - T_a, B - a)\, 1_{\{T_a \le t\}}.$$

Proof. The case $a = 0$ follows from Proposition 2.2; the case $a > 0$ and $t = 0$ is trivially true. Fix $a > 0$ and $t > 0$ and B Borel, and write T for T_a. On the event $\{T \le t\}$, we have $W_T = a$ by the continuity of W and, thus,

$$W_t = W_{T+U} - W_T + a, \qquad \text{where } U = (t - T)1_{\{T \le t\}}.$$

Hence, by Theorem 1.14 with $f = 1_{B-a}$ and, therefore, $g(u) = G(u,\; B - a)$,

$$\begin{aligned} \mathbb{E}_T 1_{\{T \le t\}}\, 1_B(W_t) &= \mathbb{E}_T\, 1_{B-a}\,(W_{T+U} - W_T)\, 1_{\{T \le t\}} \\ &= G\,(t - T, B - a)\, 1_{\{T \le t\}}. \end{aligned}$$

Taking expectations on both sides completes the proof. □

2.7 Proposition. *For a and t in $\mathbb{R}_+$, and B Borel,*

$$\mathbb{P}\{T_a \le t, W_t \in B\} = G(t, 2a - B), \qquad B \subset (-\infty, a)\,.$$

Proof. Since W is symmetric, we have $G(u, B - a) = G(u, a - B) = G(u,\; (2a - B) - a)$; thus, by the preceding lemma,

$$\mathbb{P}\{T_a \le t, W_t \in B\} = \mathbb{P}\{T_a \le t, W_t \in 2a - B\}\,.$$

If $B \subset (-\infty, a)$, then $2a - B \subset (a, \infty)$, and $W_t(\omega) > a$ implies that $T_a(\omega) \le t$. So, for $B \subset (-\infty, a)$, the right side becomes $\mathbb{P}\{W_t \in 2a - B\} = G(t,\; 2a - B)$. □

The preceding proposition is the basic computational formula. The restriction of B to subsets of $(-\infty, a)$ is without harm: we may re-state the result as

2.8 $$\mathbb{P}\{T_a > t, W_t \in B\} = G(t, B) - G(t, 2a - B), \qquad B \subset (-\infty, a),$$

and now the restriction on B is entirely logical, since the left side vanishes for subsets B of $[a, \infty)$.

In particular, taking $B = (-\infty, a)$ in 2.8, the event on the left side becomes $\{T_a > t\}$. So, since $2a - B = (a, \infty)$ then, 2.8 becomes

2.9 $$\mathbb{P}\{T_a > t\} = \mathbb{P}\{|W_t| \le a\} = 2\int_0^{a/\sqrt{t}} dx\, \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

The following collects together various interpretations of this formula.

2.10 Proposition. *Let $a > 0$. Then, $0 < T_a < \infty$ almost surely, but $\mathbb{E}T_a = +\infty$. The distribution of T_a is the same as that of a^2/Z^2, where Z is standard Gaussian. The distribution admits a continuous density function:*

2.11 $$\mathbb{P}\{T_a \in dt\} = dt\, \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}, \qquad t > 0.$$

Proof. Let Z have the standard Gaussian distribution. Then, W_t has the same distribution as $\sqrt{t}\, Z$. So, from 2.9,

$$\mathbb{P}\{T_a > t\} = \mathbb{P}\left\{\sqrt{t}\,|Z| \le a\right\} = \mathbb{P}\left\{\left(\frac{a}{Z}\right)^2 \ge t\right\} = \mathbb{P}\left\{\frac{a^2}{Z^2} > t\right\},$$

which means that T_a and a^2/Z^2 have the same distribution. Since $Z \in \mathbb{R}\backslash\{0\}$ almost surely, it follows that $T_a \in (0, \infty)$ almost surely. The density function in 2.11 is obtained by differentiating the last member of 2.9. It is seen from 2.11 that $\mathbb{E}T_a = +\infty$, since the integral of $1/\sqrt{t}$ over $(1, \infty)$ is infinity. □

The distribution in 2.11 appeared before in connection with stable processes with index 1/2; see VI.4.10 and also Chapter VII. Indeed, we shall see in the next section that $(T_a)_{a\in\mathbb{R}_+}$ is a stable Lévy process with index 1/2. For the present we note the corresponding Laplace transform (see Exercise 2.23 for one method, and 3.9 for a painless computation):

2.12 $$\mathbb{E}\, e^{-pT_a} = e^{-a\sqrt{2p}}, \qquad p \in \mathbb{R}_+.$$

Hitting times of points

The preceding Laplace transform appeared earlier, in Proposition V.5.20, for the time of entrance to $[a, \infty)$. The following is the reason for coincidence.

2.13 PROPOSITION. *Fix a in $(0, \infty)$; define*

$$T_{a-} = \inf\{t > 0 : W_t \geq a\} = \inf\{t > 0 : W_t = a\}.$$

Then, T_{a-} is a stopping time of $\mathcal{G}^o$, and $T_{a-} = T_a$ almost surely.

Proof. Write T for T_{a-}. It is obviously a $\mathcal{G}^o$-stopping time. Clearly, $T \leq T_a$. By Proposition 2.10, $T_a < \infty$ almost surely. Thus, $T < \infty$ almost surely, and $W \circ \theta_T$ is again Wiener by the strong Markov property at T. Thus, by Proposition 2.2, we have $T_0 \circ \theta_T = 0$ almost surely, which completes the proof since $T_a = T + T_0 \circ \theta_T$. □

Indeed, as the notation indicates, T_{a-} is the left-limit at a of the increasing process $b \mapsto T_b$. To see this, let (a_n) be a strictly increasing sequence with limit a. For each n, then, $T_{a_n} < \infty$ almost surely and W is at the point a_n at time T_{a_n}. Since W is continuous, it must be at the point a at the time $T = \lim T_{a_n}$. So, the limit T is equal to T_{a-}.

Hitting times of negative points

All the results above extend, by the symmetry of W, to hitting times of $(-\infty, a)$ with negative a:

2.14 $$T_a = \inf\{t > 0 : W_t < a\}, \qquad a \leq 0.$$

For $a = 0$, the hitting times of $(0, \infty)$ and $(-\infty, 0)$ are both equal to 0 almost surely, and T_0 acquires an unambiguous double-meaning.

By the symmetry of W, each T_a has the same distribution as $T_{|a|}$. Thus, T_a has the same distribution as a^2/Z^2, where Z is standard Gaussian; this is for every a in $\mathbb{R}$.

Arcsine laws

We recall some elementary facts. Let X and Y be independent standard Gaussian variables. Then, X^2 and Y^2 are independent gamma distributed with shape index ½ and scale index ½. It follows that $A = X^2/(X^2 + Y^2)$ has the beta distribution with index pair (½, ½). This particular beta is called the *arcsine distribution*, because

2.15 $$\mathbb{P}\{A \leq u\} = \int_0^u dv \frac{1}{\pi\sqrt{v(1-v)}} = \frac{2}{\pi}\arcsin\sqrt{u}, \qquad 0 \leq u \leq 1.$$

Since $C = Y/X$ has the Cauchy distribution, we also have the connection to Cauchy distribution via $A = 1/(1 + C^2)$. Another connection can be noted by recalling that C has the same distribution as $\tan B$, where the angle B has the uniform distribution on $(0, 2\pi)$; thus, $A = (\sin B)^2$ where B is uniform on $(0, 2\pi)$, which explains 2.15 above.

The following arcsine law for the Wiener process is about the probability that W does not touch 0 during the time interval $[s, u]$. A more interesting arcsine law will be given later as Theorem 6.22. For $0 \le s < u < \infty$,

2.16 $$\mathbb{P}\{W_t \in \mathbb{R}\setminus\{0\} \ \text{for all}\ t\ \in [s,u]\} = \frac{2}{\pi}\ \arcsin\sqrt{\frac{s}{u}}.$$

We shall show this as a consequence of results on the recurrence times for the point 0; see Remark 2.22 below.

Backward and forward recurrence times

Thinking of the Wiener particle, let G_t be the last time before t, and D_t the first time after t, that the particle is at the origin: for t in $\mathbb{R}_+$,

2.17 $$G_t = \sup\{s \in [0,t] : W_s = 0\}\,,\ \ D_t = \inf\{u \in (t,\infty) : W_u = 0\}\,.$$

For $t > 0$ fixed, W_t differs from 0 almost surely, which implies that $G_t < t < D_t$ almost surely. Also, in view of Corollaries 2.3 and 2.4 on the zeros of W for small and large times, it is evident that $0 < G_t$ and $D_t < \infty$ almost surely. Finally, note that D_t is a stopping time, but G_t is not; see Exercise 2.27 also.

2.18 Proposition. *Let A have the arcsine distribution as in 2.15. For each t in $\mathbb{R}_+$, then, G_t has the same distribution as tA, and D_t has the same distribution as t/A.*

Proof. Let X and Y be independent standard Gaussian variables. Recall that $T_a \approx a^2/Y^2$ for every a, where the symbol "$\approx$" stands for "has the same distribution as".

Consider $R_t = D_t - t$. If $W_t(\omega) = x$, then $R_t(\omega)$ is the hitting time of the point $-x$ by the path $W(\theta_t\omega)$. Since $W \circ \theta_t$ is Wiener independent of $\mathcal{F}_t$, and similarly for $(-W)\circ\theta_t$ by symmetry, we conclude that $R_t \approx {W_t}^2/Y^2$, where Y is independent of W_t. Thus, we may replace W_t with $\sqrt{t}X$; we obtain $R_t \approx t\,X^2/Y^2$. Hence,

$$D_t = t + R_t \approx t\left(X^2 + Y^2\right)/Y^2 \approx t/A$$

as claimed. Finally, $G_t \approx t\,A$ since, for s in $(0,\,t)$,

$$\mathbb{P}\{G_t < s\} = \mathbb{P}\{D_s > t\} = \mathbb{P}\left\{\frac{s}{A} > t\right\} = \mathbb{P}\{tA < s\}\,. \qquad \square$$

The terms forward and backward recurrence times refer to the variables

2.19 $$R_t = D_t - t, \qquad Q_t = t - G_t.$$

within the proof, it is shown that $R_t \approx t\,X^2/Y^2 = t\,C^2$, where C has the standard Cauchy distribution. The distribution of Q_t is the same as that of G_t:

2.20 $$\mathbb{P}\{G_t \le s\} = \mathbb{P}\{Q_t \le s\} = \frac{2}{\pi}\ \arcsin\sqrt{\frac{s}{t}}, \quad 0 \le s \le t;$$

this is because A and $1 - A$ have the same distribution. Various joint distributions can be obtained from the observation that

2.21 $$\{G_u < s\} = \{D_s > u\} = \{G_t < s, D_t > u\}, \qquad 0 \le s < t < u.$$

We put some such as exercises.

2.22 REMARK. Arcsine law 2.16 is a consequence of the arcsine distribution for G_t, because the event on the left side of 2.16 is the same as $\{G_u < s\}$.

Exercises

2.23 *Laplace transform for T_a.* This is to avoid a direct computation using the distribution 2.11. First, use 2.9 to show that

$$\mathbb{E}\, e^{-pT_a} = \int_0^\infty dt\; pe^{-pt}\; \mathbb{P}\{T_a \le t\} = \mathbb{P}\{|W_S| > a\}, \quad p \ge 0,$$

where S is independent of W and has the exponential distribution with parameter p. Recall that, then, W_S has the same distribution as $S_1 - S_2$, where S_1 and S_2 are independent exponential variables with parameter $\sqrt{2p}$. Conclude that 2.12 holds.

2.24 *Potentials.* Let $X = X_0 + W$ be the standard Brownian motion with initial state X_0. Write $\mathbb{E}^x$ for the expectation operator given that $X_0 = x$. For Borel $f : \mathbb{R} \mapsto \mathbb{R}_+$, define

$$\begin{aligned} U_p f(x) &= \mathbb{E}^x \int_0^\infty dt\; e^{-pt}\; f \circ X_t \\ &= \mathbb{E} \int_0^\infty dt\; e^{-pt}\; f(x + W_t), \qquad p \in \mathbb{R}_+,\ x \in \mathbb{R}. \end{aligned}$$

The function $U_p f$ is called the p-potential of f. Show that, for $p > 0$,

$$U_p f(x) = \int_{\mathbb{R}} dy\; u_p(x - y)\; f(y),$$

where

$$u_p(x) = \int_0^\infty dt\; e^{-pt}\; \frac{e^{-x^2/2t}}{\sqrt{2\pi t}} = \frac{1}{\sqrt{2p}}\; e^{-\sqrt{2px^2}}, \; x \in \mathbb{R}.$$

2.25 *Zeros to left and right.* With G_t and D_t defined by 2.17, show that, for $0 < s < t < u$,

$$\mathbb{P}\{G_t \in ds\} = ds\; \tfrac{1}{\pi\sqrt{s(t-s)}}, \quad \mathbb{P}\{D_t \in du\} = du\; \tfrac{t}{\pi u\sqrt{t(u-t)}},$$
$$\mathbb{P}\{G_t \in ds,\; D_t \in du\} = ds\; du\; \tfrac{1}{2\pi\sqrt{s(u-s)^3}}.$$

2.26 *Recurrence times.* For R_t and Q_t defined by 2.19, show that, for $0 < q < t$ and $r \geq 0$,

$$\mathbb{P}\{Q_t \in dq\} = dq\ \frac{1}{\pi\sqrt{q(t-q)}},\ \mathbb{P}\{R_t \in dr | Q_t = q\} = \frac{1}{2}\sqrt{\frac{q}{(q+r)^3}}.$$

2.27 *No stopping at G_t.* Of course, G_t is not a stopping time. This is to show that, moreover, G_t has no chance of coinciding with a stopping time: Let S be a stopping time of $\mathcal{F}$. We shall show that

$$\mathbb{P}\{S = G_t\} = 0.$$

By replacing S with $S \wedge t$, we may assume that $S \leq t$.

a) Show that $T_0 \circ \theta_S = 0$ almost surely; this is by the strong Markov property coupled with Proposition 2.2.

b) Show that, for almost every ω and every $\varepsilon > 0$, there is u in the interval $(S(\omega),\ S(\omega) + \varepsilon)$ such that $W_u(\omega) = 0$.

c) Show that the preceding statement is incompatible with the definition of G_t for ω in $\{S = G_t\}$.

3 Hitting Times and Running Maximum

The setup is as in the preceding section. We are interested in the process $T = (T_a)_{a\in\mathbb{R}_+}$ of hitting times and its relationship to the process $M = (M_t)_{t\in\mathbb{R}_+}$ of *running maximum*, where

3.1 $$M_t(\omega) = \max\nolimits_{0\leq s\leq t} W_s(\omega), \qquad t \in \mathbb{R}_+,\ \omega \in \Omega.$$

The definition 2.1 of $T_a(\omega)$ remains true when $W_t(\omega)$ there is replaced with $M_t(\omega)$. Indeed, the paths $a \mapsto T_a(\omega)$ and $t \mapsto M_t(\omega)$ are functional inverses of each other:

3.2 $$T_a(\omega) = \inf\{t > 0 : M_t(\omega) > a\},\ M_t(\omega) = \inf\{a > 0 : T_a(\omega) > t\}.$$

This relationship, together with the previous results on the T_a, shows that the following holds; see Figure 11 below as well. No further proof seems needed.

3.3 Lemma. *For almost every ω, the path $a \mapsto T_a(\omega)$ is right-continuous, strictly increasing, real-valued, and with $T_o(\omega) = 0$ and $\lim_{a\to\infty} T_a(\omega) = +\infty$. For almost every ω, the path $t \mapsto M_t(\omega)$ is increasing, continuous, real-valued, and with $M_o(\omega) = 0$ and $\lim_{t\to\infty} M_t(\omega) = +\infty$.*

In particular, $T_a(\omega) < t$ if and only if $M_t(\omega) > a$, this being true for every a and t in $\mathbb{R}_+$. Thus, the formula 2.9 may be re-stated as follows.

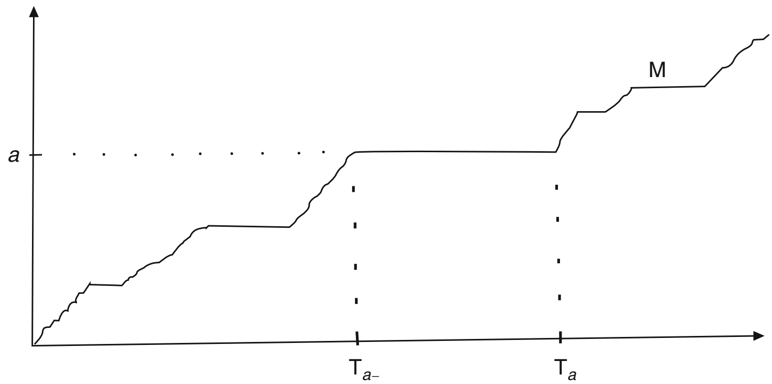


Figure 11: The path M is increasing and continuous; T_a is the time it hits the interval (a, ∞).

3.4 PROPOSITION. *For every a and t in $\mathbb{R}_+$,*

$$\mathbb{P}\{T_a < t\} = \mathbb{P}\{M_t > a\} = \mathbb{P}\{|W_t| > a\}.$$

3.5 REMARK. The preceding implies that M_t has the same distribution as $|W_t|$ for each t; thus, $\mathbb{E}\, M_t = \sqrt{2t/\pi}$ and $\mathbb{E}\, M_t^2 = t$ in particular. The probability law of the process M, however, is very different from that of $|W|$. The law of M is specified by the relationship 3.2 and the law of the process (T_a).

Hitting time process is stable Lévy

3.6 THEOREM. *The process $T = (T_a)_{a\in\mathbb{R}_+}$ is a strictly increasing pure-jump Lévy process. It is stable with index 1/2, and its Lévy measure is*

3.7 $$\lambda(dt) = dt\frac{1}{\sqrt{2\pi t^3}}, \qquad t > 0.$$

Proof. Fix a and b in $(0, \infty)$. In order for the process W to hit the interval $(a+b, \infty)$, it must hit (a, ∞) first, and, then, the future process $W \circ \theta_{T_a}$ must hit (b, ∞); in short,

$$T_{a+b} = T_a + T_b \circ \theta_{T_a}.$$

Since $T_a < \infty$ almost surely, the process $W \circ \theta_{T_a}$ is independent of $\mathcal{F}_{T_a}$ and is again a Wiener process; this is by the strong Marker property at T_a. Thus, $T_{a+b} - T_a = T_b \circ \theta_{T_a}$ is independent of $\mathcal{F}_{T_a}$ and has the same distribution as T_b. Together with Lemma 3.3, this shows that the process T is a strictly increasing Lévy process over the stochastic base $(\Omega, \mathcal{H}, \hat{\mathcal{F}}, \hat{\theta}, \mathbb{P})$, where $\hat{\mathcal{F}}_a = \mathcal{F}_{T_a}$ and $\hat{\theta}_a = \theta_{T_a}$; see Definition VII.3.3.

The distribution of T_a is the same as that of $a^2 T_1$; this is by Proposition 2.10. Thus, the Lévy process T is stable with index 1/2. Every such process is of the pure-jump type, and its Lévy measure has the form $\lambda(dt) = dt\ c/t^{3/2}$; see Example VII.2.1. Finally, the constant c must be equal to $1/\sqrt{2\pi}$ in this case, since VII.2.1 and 2.12 imply

$$\mathbb{E}\ e^{-pT_a} = \exp_-\ a \int_{\mathbb{R}_+} \lambda(dt)\left(1 - e^{-pt}\right) = \exp_-\ a\sqrt{2p}.$$

□

Poisson jump measure

We use the preceding theorem to clarify the fine structure of the processes T and M. Recall the Itô-Lévy decomposition for Lévy processes; see Theorem VII.5.2 and VII.5.14 and *et seq.* The following needs no further proof.

3.8 THEOREM. *Let N be the random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ defined by*

$$N(\omega, B) = \sum_a 1_B\left(a, T_a(\omega) - T_{a-}(\omega)\right), \qquad \omega \in \Omega,\ B \in \mathcal{B}\left(\mathbb{R}_+ \times \mathbb{R}_+\right),$$

where the sum is over all a for which $T_a(\omega) > T_{a-}(\omega)$. Then, N is Poisson with mean measure Leb $\times\ \lambda$*, where λ is as given by 3.7. Conversely,*

$$T_a(\omega) = \int_{(0,a]\times\mathbb{R}_+} N(\omega; db, du)\ u, \qquad a \in \mathbb{R}_+,\ \omega \in \Omega.$$

The relationship between the random measure N and the processes M and T are shown in the Figure 12 below. We describe some of the features:

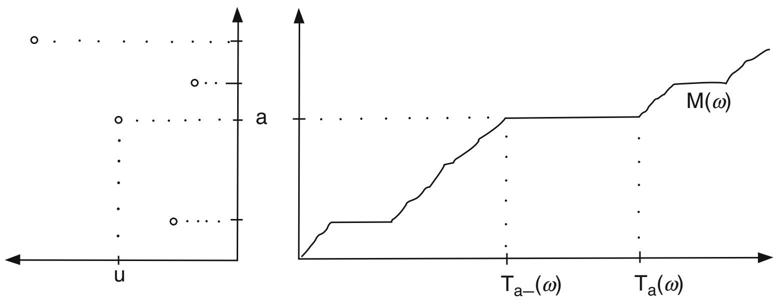


Figure 12: Big sized atoms of $N(\omega, \cdot)$ are marked with little circles on the graph left. Corresponding to the atom (a, u), there is a jump of size u from $T_{a-}(\omega)$ to $T_{a-}(\omega) + u = T_a(\omega)$, the path $M(\omega)$ stays constant at level a during the time interval $[T_{a-}(\omega),\ T_a(\omega)]$.

The following holds for almost every ω: A point (a, u) is an atom of the counting measure $N(\omega, \cdot)$ if and only if the path $M(\omega)$ has a flat stretch of length u at the level a, and then, the hitting time $T_a(\omega)$ of the interval (a, ∞) is exactly u time units later than the hitting time $T_{a-}(\omega)$ of the point a. Since $N(\omega, \cdot)$ has only countably many atoms, this situation occurs at countably many a only. Since there are infinitely many atoms in the set $(a, b) \times (0, \infty)$, the path $M(\omega)$ stays flat at infinitely many levels on its way from a to b; however, for $\varepsilon > 0$ however small, only finitely many of those sojourns exceed ε in duration.

The situation at a fixed level a is simpler. For $a > 0$ fixed, almost surely, there are no atoms on the line $\{a\} \times \mathbb{R}_+$; therefore, $T_a = T_{a-}$ almost surely.

Exercises

3.9 *Time change.* Show that, for every p in $\mathbb{R}_+$,

$$\int_{\mathbb{R}_+} e^{-pt}\, dM_t = \int_{\mathbb{R}_+} da\ e^{-pT_a}.$$

This suggests a painless way of computing the Laplace transform for T_a. Since (T_a) is Lévy, the Laplace transform has the form $e^{-a\varphi(p)}$. Hence, the expected value of the right side above is equal to $1/\varphi(p)$. Whereas, the expected value of the left side is easy to compute using $\mathbb{E}\, M_t = \sqrt{2t/\pi}$; the result is $1/\sqrt{2p}$. So, $\varphi(p) = \sqrt{2p}$, confirming 2.12 once more.

3.10 *Cauchy connection.* Let X be a Wiener process independent of W and, thus, independent of (T_a). Show that $(X_{T_a})_{a \in \mathbb{R}_+}$ is a Cauchy process; see Example VII.2.14 for Cauchy.

3.11 *Continuation.* Let $(X_t,\ Y_t)_{t \in \mathbb{R}_+}$ be a standard Brownian motion in $\mathbb{R}^2$ with initial state $(X_0,\ Y_0) = (0,\ y)$ for some fixed $y > 0$. Let S be the first time that the motion (X, Y) touches the x-axis. Find the distribution of X_S, the point touched on the x-axis.

4 Wiener and its Maximum

The setup and notations are as before in Sections 2 and 3. Our aim is to examine the joint law of the Wiener process W and its running maximum M defined by 3.1. We shall see that $M - W$ is a reflected Brownian motion and that it determines both M and W. As a supplement, we mention that $2M - W$ is a Bessel process of index 3, that is, it has the same law as the radial Brownian motion in dimension 3. These results will lead to excursions and local times in the next sections.

Distribution of M and W at a fixed time

4.1 PROPOSITION. *For fixed times $t > 0$,*

$$\mathbb{P}\{M_t \in da,\ M_t - W_t \in db\} = da\ db\ \frac{2(a+b)e^{-(a+b)^2/2t}}{\sqrt{2\pi t^3}},\ a, b \in \mathbb{R}_+.$$

Proof. Recall that $T_a(\omega) < t$ if and only if $M_t(\omega) > a$, and that the distribution of T_a is diffuse. Thus, we may re-write Proposition 2.7 in the form

$$\mathbb{P}\{M_t > a, W_t \le x\} = \mathbb{P}\{W_t > 2a - x\} = \int_{2a-x}^{\infty} dy\ \frac{e^{-y^2/2t}}{\sqrt{2\pi t}},\ x \le a,$$

Differentiating this with respect to a and x, and putting $a - x = b$, we see that the claimed expression holds. □

In the preceding proposition, it is worth noting the symmetry with respect to the arguments a and b. It follows that $M_t - W_t$ and M_t have the same marginal distribution, and the distribution of the latter is the same as that of $|W_t|$; see 3.4. This proves the following.

4.2 COROLLARY. *For fixed t, the variables M_t, $|W_t|$, and $M_t - W_t$ have the same distribution.*

As a process, M is very different from $|W|$ and $M - W$. But, the latter two are alike: they have the same law; see 4.6 below.

Construction of M from the zeros of $M - W$

Fix an outcome ω. The set

$$D_\omega = \{t \in \mathbb{R}_+ : \quad M_t(\omega) - W_t(\omega) > 0\} \tag{4.3}$$

is open, since it is the inverse image of the open set $(0, \infty)$ under the continuous mapping $t \mapsto M_t(\omega) - W_t(\omega)$. Thus, D_ω is a countable union of disjoint open intervals. For $\varepsilon > 0$, let $N_t(\omega, \varepsilon)$ be the number of those open intervals contained in $[0, t]$ and having lengths exceeding ε.

4.4 THEOREM. *For almost every ω,*

$$\lim_{\varepsilon \downarrow 0} \sqrt{2\pi\varepsilon}\ N_t(\omega, \varepsilon) = 2M_t(\omega), \quad t \in \mathbb{R}_+.$$

REMARK. This shows that M is determined by N, which is in turn determined by the zero-set of $M - W$. Interestingly, thus, $M - W$ determines both M and W.

Proof. In terms of the Poisson random measure N described by Theorem 3.8,

$$N_t(\omega, \varepsilon) = N(\omega, (0, M_t(\omega)) \times (\varepsilon, \infty)).$$

Thus, it is sufficient to show that, for each a in $\mathbb{R}_+$, almost surely,

4.5 $$\lim_{\varepsilon \downarrow 0} \sqrt{2\pi\varepsilon}\ N\left((0, a) \times (\varepsilon, \infty)\right) = 2a.$$

Recalling the mean measure of N (see 3.7 and 3.8), we have

$$\mathbb{E}\ N\left((0, a) \times \left(\frac{1}{k^2}, \infty\right)\right) = a \int_{1/k^2}^{\infty} dt \frac{1}{\sqrt{2\pi t^3}} = \frac{2\ a}{\sqrt{2\pi}}\ k.$$

Thus, since N is Poisson, the right side of the expression

$$N\left((0, a) \times \left(\frac{1}{n^2}, \infty\right)\right) = \sum_{k=1}^{n} N\left((0, a) \times \left(\frac{1}{k^2}, \frac{1}{(k-1)^2}\right]\right)$$

is the sum of n independent and identically distributed random variables with mean $2a/\sqrt{2\pi}$ each. Hence, by the strong law of large numbers,

$$\lim_{n \to \infty} \frac{1}{n} N\left((0, a) \times \left(\frac{1}{n^2}, \infty\right)\right) = \frac{2a}{\sqrt{2\pi}}$$

almost surely. This proves 4.5 and completes the proof of 4.4. □

The preceding theorem can be strengthened: For almost every ω, the convergence shown is indeed uniform in t over compacts.

Process $M - W$ is a reflected Wiener

The following is the main result of this section. We shall prove it by constructing a Wiener process V such that $|V| = M - W$.

4.6 Theorem. *The processes $M - W$ and $|W|$ have the same law.*

4.7 Remark. This theorem is a corollary to Theorem 4.8 below, where we show the existence of a Wiener process V such that $M - W = |V|$. We start by analyzing the problem of constructing V.

Observing $M_t(\omega)$ and $W_t(\omega)$ yields only the absolute value $|V_t(\omega)|$; to obtain $V_t(\omega)$ we need to supply the sign. To see how this should be done, we examine Figure 13 below. Note that the path $W(\omega)$ coincides with the path $M(\omega)$ at all times except those belonging to the open set D_ω defined

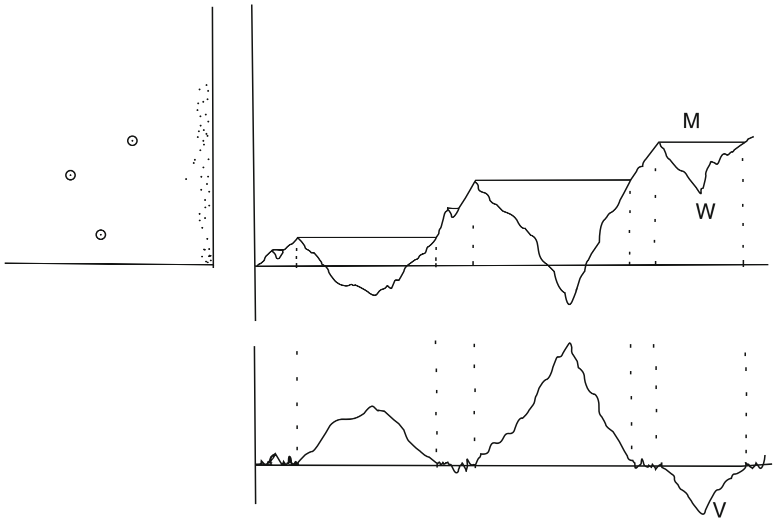


Figure 13: The path $M(\omega)$ is increasing continuous; $W(\omega)$ hangs like a stalactite from each flat stretch of $M(\omega)$. To construct $V(\omega)$, each stalactite is made to stand up or hang down from the time axis, the two choices being equally likely.

by 4.3. Over each component of D_ω, the path $M(\omega)$ stays flat and the path $W(\omega)$ hangs from $M(\omega)$ like a stalactite. Over the same interval, then, $V(\omega)$ will have to be either a stalactite hanging from the time axis, or a stalagmite standing up, the two possibilities being equally likely. Thus, we need to assign a sign, either positive or negative, to each stalactite hanging from $M(\omega)$.

To provide the needed signs, we need, independent of W, a countable independency of Bernoulli variables taking the values $+1$ and -1 with equal probabilities. If $(\Omega, \mathcal{H}, \mathbb{P})$ does not support such a sequence $(B_i)_{i\in\mathbb{N}}$, we enlarge it as follows: Let $D = \{+1, -1\}$, $\mathcal{D} = 2^D$, $\mu = \frac{1}{2}\,\delta_1 + \frac{1}{2}\,\delta_{-1}$ with δ_x being Dirac at x as usual; replace $(\Omega, \mathcal{H}, \mathbb{P})$ with

$$\left(\hat{\Omega}, \hat{\mathcal{H}}, \hat{\mathbb{P}}\right) = (\Omega, \mathcal{H}, \mathbb{P}) \times (D, \mathcal{D}, \mu)^{\mathbb{N}},$$

and, for $\hat{\omega} = (\omega, \omega')$ in $\hat{\Omega}$, define $\hat{W}_t(\hat{\omega}) = W_t(\omega)$ and let $B_i(\hat{\omega})$ be the i-coordinate of ω'. In the next theorem, we shall assume that this enlargement, if needed, is done already. Theorem 4.6 is a corollary of the next theorem.

4.8 Theorem. *There exists (on a possibly enlarged probability space) a Wiener process V such that $M - W = |V|$.*

Proof. We may and do assume that there is, independent of W, an independency $(B_i)_{i\in\mathbb{N}}$ of variables taking the values $+1$ and -1 with equal probabilities.

a) Let N be the Poisson random measure described by Theorem 3.8, and let $(A_i, U_i)_{i\in\mathbb{N}}$ be a labeling of its atoms. Then, the triplets (A_i, U_i, B_i) are the atoms of a Poisson random measure $\hat{N}$; see Corollary VI.3.5.

Fix ω, and let (a, u, b) be an atom of $\hat{N}(\omega, \cdot)$. Corresponding to that atom, $M(\omega)$ remains equal to a over the time interval $(s, s+u) = (T_{a-}(\omega),\ T_a(\omega))$; we define

4.9 $$V_t(\omega) = (M_t(\omega) - W_t(\omega))\, b, \qquad t \in (s, s+u)\,.$$

Doing this for every atom, we obtain $V_t(\omega)$ for every t for which $M_t(\omega) \neq W_t(\omega)$; for all other t, we define $V_t(\omega) = M_t(\omega) - W_t(\omega) = 0$.

For fixed t, we remarked in Corollary 4.2 that $M_t - W_t$ has the same distribution as $|W_t|$. Thus, in view of 4.9 and the independence of the Bernoulli variables B_i from W,

4.10 $$\mathbb{P}\{V_t \in A\} = \mathbb{P}\{W_t \in A\} = G\,(t, A)\,, \qquad A \in \mathcal{B}_{\mathbb{R}},$$

with the same notation 2.5 for the Gaussian kernel G.

b) It is obvious that V is continuous and starts from $V_0 = 0$. To show that it is Wiener, we shall show that

4.11 $$\mathbb{P}\left\{V_{s+t} - V_s \in A \,|\, \hat{\mathcal{F}}_s\right\} = G(t, A), \qquad s, t \in \mathbb{R}_+,\ A \in \mathcal{B}_{\mathbb{R}},$$

where $\hat{\mathcal{F}}_s$ is the σ-algebra generated by the union of $\mathcal{F}_s$ and $\sigma\{V_r : r \le s\}$. This is obvious if $s = 0$ or $t = 0$. For the remainder of the proof, we fix $s > 0$ and $t > 0$ and define

$$D = \inf\{u > s : W_u = M_s\}\,, \quad R = D - s.$$

Observe that D is a stopping time of $\mathcal{F}$ and, thus, of $\hat{\mathcal{F}}$; moreover, almost surely, $D < \infty,\ W_D = M_D = M_s,\ V_D = 0$. It is clear that, in view of 4.10,

4.12 $$\mathbb{P}\{V_{D+u} - V_D \in A\} = \mathbb{P}\{V_u \in A\} = G(u, A).$$

c) On the event $\{R \le t,\ V_s = x\}$, we have $s < D \le s+t$ and $V_D = 0$ and

$$V_{s+t} = V_{D+(t-R)} - V_D.$$

Thus, as in Theorem 1.14, it follows from 4.12 that

$$\mathbb{P}\left\{R \le t, V_{s+t} - V_s \in A \,|\, \hat{\mathcal{F}}_D\right\} = G(t - R, A + x)1_{\{R \le t\}}$$

on $\{V_s = x\}$, on which we also have $R = T_a \circ \theta_s$ with $a = |x|$. Hence, conditioning both sides on $\hat{\mathcal{F}}_s$, since $\hat{\mathcal{F}}_s \subset \hat{\mathcal{F}}_D$ and $T_a \circ \theta_s$ is independent of $\hat{\mathcal{F}}_s$ and has the same distribution as T_a, we get

4.13 $$\mathbb{P}\left\{R \le t, V_{s+t} - V_s \in A \,|\, \hat{\mathcal{F}}_s\right\} = f \circ V_s$$

where

4.14 $$f(x) = \mathbb{E}\, G\,(t - T_a, A + x)\, 1_{\{T_a \le t\}}, \quad x \in \mathbb{R},\ a = |x|\,.$$

d) On $\{R > t,\ V_s = x\}$, the variable V_{s+t} has the same sign as x, and

$$R = T_a \circ \theta_s,\ V_{s+t} - V_s = -b\ W_t \circ \theta_s \quad \text{with} \quad a = |x|\,,\ b = \text{sgn}\ x.$$

Thus, by the Markov property of W,

4.15 $$\mathbb{P}\left\{R > t, V_{s+t} - V_s \in A \,|\, \hat{\mathfrak{F}}_s\right\} = g \circ V_s,$$

where

$$g(x) = \mathbb{P}\{T_a > t, \quad -bW_t \in A\}\,, \qquad x \in \mathbb{R},\ a = |x|,\ b = \text{sgn}\ x.$$

We use Lemma 2.6 and the symmetry of $G(u, \cdot)$ to evaluate $g(x)$:

4.16 $$\begin{aligned} g(x) &= \mathbb{P}\{W_t \in -bA\} - \mathbb{P}\{T_a \le t, W_t \in -bA\} \\ &= G(t, A) - \mathbb{E}\, G(t - T_a, -bA - a) 1_{\{T_a \le t\}} \\ &= G(t, A) - \mathbb{E}\, G(t - T_a, A + x) 1_{\{T_a \le t\}}, \end{aligned}$$

where the last equality is justified by noting that $-bA - a$ is equal to $-A - x$ if $x \ge 0$ and to $A + x$ if $x \le 0$.

e) It follows from 4.14 and 4.16 that $f(x) + g(x) = G(t,\ A)$, and we obtain 4.11 by putting 4.13 and 4.15 together. □

Process $2M - W$ is a Bessel with index $d = 3$

Let $R = 2M - W$. Since $M - W \ge 0$, we have $R \ge M$. Recalling that M is increasing and strictly positive on $(0, \infty)$ and with limit equal to $+\infty$ as $t \to \infty$, we conclude the following: For almost every ω, we have

4.17 $\quad R_0(\omega) = 0, \quad R_t(\omega) > 0 \quad$ for every $t > 0,\ \lim_{t\to\infty} R_t(\omega) = +\infty.$

For each ω, the path $R(\omega)$ is obtained by reflecting the path $W(\omega)$ at its running maximum $M(\omega)$, that is, each stalactite of $W(\omega)$ hanging from $M(\omega)$ is made into a stalagmite sitting on $M(\omega)$. From this picture, it is now evident that

4.18 $$M_t(\omega) = \inf_{u \ge t}\ R_u(\omega), \qquad t \in \mathbb{R}_+.$$

Thus, the path $R(\omega)$ defines the path $M(\omega)$ and, hence, the path $W(\omega) = 2M(\omega) - R(\omega)$.

Recall from Example 1.22 that a Bessel process of index $d = 3$ is a Markov process whose law is identical to that of $|X|$, where X is a 3-dimensional Wiener process. The proof of the following will be sketched in Exercises 4.27.

4.19 THEOREM. *The process $R = 2M - W$ is a Bessel process of index $d = 3$.*

The preceding clarifies the recurrence properties of Wiener processes in $\mathbb{R}^d$ for $d \geq 3$. Let X be a Wiener process in $\mathbb{R}^3$. According to the preceding theorem, its radial part $|X|$ has the same law as $R = 2M - W$. It follows from 4.17 that, for almost every ω,

4.20 $$|X_0(\omega)| = 0, \;\; |X_t(\omega)| > 0 \text{ for every } t, \;\; \lim_{t\to\infty} |X_t(\omega)| = +\infty.$$

Thus, the Wiener particle X in $\mathbb{R}^3$ starts from the origin, and never returns to the origin, and the set of times spent in a bounded Borel set B is bounded. The process X is transient in this sense. The same statements are true for a Wiener process X in $\mathbb{R}^d$ with $d > 3$, since every choice of three components of X define a Wiener process in $\mathbb{R}^3$.

Exercises

4.21 *Arcsine law for M.*

a) Fix t. Show that, for almost every ω, $W_t(\omega) = M_t(\omega)$ if and only if $M_{t+\varepsilon}(\omega) > M_t(\omega)$ for every $\varepsilon > 0$.

b) Show that, for $0 < s < t$, the event $\{M_s = M_t\}$ and the event $\{W_u < M_u, \; s < u < t\}$ have the same probability.

c) Show that $\mathbb{P}\{M_s = M_t\} = \frac{2}{\pi} \arcsin \sqrt{s/t}$.

4.22 *Continuation.* For $t > 0$, let $\hat{G}_t = \sup\{s \leq t : W_s = M_s\}$. Compute the distribution of $\hat{G}_t$.

4.23 *Some joint distributions.* It will be convenient to introduce

$$h_t(a) = \frac{a\; e^{-a^2/2t}}{\sqrt{2\pi t^3}}, \qquad k_t(x,y) = \frac{e^{-(x-y)^2/2t}}{\sqrt{2\pi t}} - \frac{e^{-(x+y)^2/2t}}{\sqrt{2\pi t}}$$

for $t > 0$, $a > 0$, and x and y real. Note that $t \mapsto h_t(a)$ is the density for the distribution of T_a, the hitting time of a; thus

$$\int_0^t ds \quad h_s(a)\; h_{t-s}(b) = h_t(a+b), \qquad a, b > 0.$$

a) Show that, for a and b in $(0,\infty)$,

$$\int_0^t ds\; \frac{a\; e^{-a^2/2s}}{\sqrt{2\pi s^3}} \cdot \frac{e^{-b^2/2(t-s)}}{\sqrt{2\pi(t-s)}} = \frac{e^{-(a+b)^2/2t}}{\sqrt{2\pi t}}.$$

b) Show that, for a and x in $(0, \infty)$,

$$\mathbb{P}\{T_a > t,\ W_t \in a - dx\} = dx\ k_t(a, x),$$
$$\mathbb{P}\{M_t \in da, W_t \in a - dx\} = da\ dx\ 2\ h_t(a + x).$$

c) Another interpretation for $k_t(x, y)$: show that

$$\mathbb{P}\{W_{s+t} \in dy; \quad W_u \neq 0 \text{ for } u\ \in (s, s+t) | W_s = x\} = dy\ k_t(x, y)$$

provided that x and y are either both positive or both negative.

4.24 *The process* $Y = (M,\ M - W)$. This is clearly a Markov process with state space $\mathbb{R}_+ \times \mathbb{R}_+$: for s, t in $\mathbb{R}_+$ and B in $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)$

$$\mathbb{P}\{Y_{s+t} \in B | \mathcal{F}_s\} = P_t(Y_s, B).$$

The transition kernel P_t can be computed explicitly: in terms of h_t and k_t introduced above, for a, b, y in $(0, \infty)$ and $x \geq a$,

$$\begin{aligned} &P_t\,(a, b; dx,\ dy) \\ &\quad = \mathbb{P}\{M_{s+t} \in dx, M_{s+t} - W_{s+t} \in dy \mid M_s = a, M_s - W_s = b\} \\ &\quad = \mathbb{P}\{T_b > t, b - W_t \in dy\}\, I(a, dx) \\ &\qquad + \int_0^t \mathbb{P}\{T_b \in t - du\}\ \mathbb{P}\{a + M_u \in dx, M_u - W_u \in dy\} \\ &\quad = I(a, dx)\ dy\ k_t(b, y) + dx\ dy\ 2h_t(x - a + b + y). \end{aligned}$$

4.25 *A martingale.* For fixed p in $\mathbb{R}_+$,

$$Z_t = e^{-pM_t}\left[1 + p(M_t - W_t)\right], \qquad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale. Show this via the following steps.

a) Use 4.1 to show directly that $\mathbb{E}\, Z_t = 1$ for every t.

b) In terms of the process Y of 4.24, note that $Z_t = f \circ Y_t$, where $f(x,\ y) = e^{-px}(1 + py)$ for x, y in $\mathbb{R}_+$. Use the Markov property for Y to show that Z is a martingale if $P_t f = f$, that is, if f is harmonic for Y.

c) Use part (a) here and some of the stages in 4.24 to show that

$$\begin{aligned} P_t f(a, b) &= e^{-pa}\ \mathbb{E}(1 + p(b - W_t))1_{\{T_b > t\}} \\ &\qquad + \int_0^t \mathbb{P}\{T_b \in t - du\}\ \mathbb{E}\ e^{-p(a + M_u)}(1 + p(M_u - W_u)) \\ &= e^{-pa} + pe^{-pa}\ \mathbb{E}(b - W_t)1_{\{T_b > t\}}. \end{aligned}$$

d) To conclude that $P_t f(a,\ b) = f(a,\ b)$, show that

$$\mathbb{E}(b - W_t)1_{\{T_b > t\}} = \mathbb{E}(b - W_t) + \mathbb{E}(W_t - b)1_{\{T_b \le t\}} = b + 0 = b.$$

4.26 *The process* $(M, 2M - W)$. Define $R = 2M - W$ as in Theorem 4.19. In preparation for the proof of 4.19, we consider the process $(M,\ R)$. It is obvious that $(M,\ R)$ is a Markov process whose state space is the region of $\mathbb{R}_+ \times \mathbb{R}_+$ above the diagonal. Of course, $M_0 = R_0 = 0$.

Show that, for $t > 0$,

$$\begin{aligned}
\mu_t(dx, dy) &= \mathbb{P}\{M_t \in dx, R_t \in dy\} = dx\ dy\ 2h_t(y),\ \ 0 \le x \le y. \\
Q_t(a, b; dx, dy) &= \mathbb{P}\{M_{s+t} \in dx, R_{s+t} \in dy | M_s = a, R_s = b\} \\
&= I(a, dx)\ dy\ k_t(b - a, y - x) + dx\ dy\ 2h_t(b + y - 2a)
\end{aligned}$$

for $0 < a \le b,\ a \le x \le y$. In particular, given that $R_t = y$, the conditional distribution of M_t is uniform on $(0,\ y)$.

4.27 *Proof of Theorem* 4.19. For the process $R = 2M - W$, the results of the preceding exercise can be used to compute that

$$\begin{aligned}
\nu_t(dx) &= \mathbb{P}\{R_t \in dx\} = dx\ 2\ x\ h_t(x) \\
P_t(x, dy) &= \mathbb{P}\{R_{s+t} \in dy | R_s = x\} = dy\ \frac{y}{x}\ k_t(x, y)
\end{aligned}$$

for $t > 0$ and $x,\ y > 0$, of course, $R_0 = 0$. These results coincide with their counterparts in Example 1.22 (see 1.25 and 1.26) for the Bessel process with index $d = 3$. To show that $R = 2M - W$ is a Bessel process with index 3, there remains to show that R is a Markov process. There does not seem to be an elegant proof. A direct proof, elementary but computationally intensive, can be obtained as follows.

Fix an integer $n \ge 2$, and a positive Borel function on $\mathbb{R}_+^n$. For times $0 < t_1 < t_2 < \ldots < t_n$, by the Markov property of (M, R), we have

$$\begin{aligned}
&\mathbb{E} f(R_{t_1}, \ldots, R_{t_n}) \\
&\quad = \int \mu_{t_1}(dx_1, dy_1) \int Q_{t_2 - t_1}(x_1, y_1; dx_2, dy_2) \int \cdots \\
&\qquad\quad \int Q_{t_n - t_{n-1}}(x_{n-1}, y_{n-1}; dx_n, dy_n)\, f(y_1, y_2, \ldots, y_n)
\end{aligned}$$

We need to show that the right side is as it should be, that is, that the right side is equal to

$$\int \nu_{t_1}(dy_1) \int P_{t_2 - t_1}(y_1, dy_2) \int \cdots \int P_{t_n - t_{n-1}}(y_{n-1}, dy_n)\, f(y_1, \ldots, y_n)\,.$$

5 Zeros, Local Times

We keep the setup and notations of the previous sections: $W = (W_t)$ is a Wiener process, $M = (M_t)$ is its running maximum, and $T = (T_a)$ is the process of hitting times. We are interested in the Cantor set like features of the set C of times at which W is at 0, and in the existence of a random measure whose support is C, called the *local time* measure.

Closed and perfect sets

This is to review some terminology. Let C be a closed subset of $\mathbb{R}_+$. Then, its complement $\mathbb{R}_+\backslash C$ is open and, therefore, is a countable union of disjoint open intervals. Those open intervals are said to be *contiguous* to C. A point of C is *isolated* if it is the common end point of two distinct contiguous intervals, or, if it is zero and is the left-end point of a contiguous interval. The set C is *dense in itself* if it has no isolated points, that is, if every point of C is a limit point of C.

A *perfect* set is a closed set with no isolated points. The simplest example is a union of finitely many disjoint closed intervals. Another example, closer to our present concerns, is the Cantor set. Every perfect set has the power of the continuum, that is, there exists an injection of $\mathbb{R}_+$ into C; see I.5.22 for this with the Cantor set.

Zeros of W

We are interested in the qualitative features of the set

5.1 $$C_\omega = \{t \in \mathbb{R}_+ : W_t(\omega) = 0\}, \quad \omega \in \Omega,$$

the set of zeros of W. For fixed ω, it is the inverse image of the closed set $\{0\}$ under the continuous mapping $t \mapsto W_t(\omega)$ from $\mathbb{R}_+$ into $\mathbb{R}$; thus, it is closed, and its complement is the union of a countable collection of disjoint open intervals, called *contiguous intervals.*

Fix the integers m and n in $\mathbb{N}^*$. Consider those intervals contiguous to C_ω whose lengths belong to the interval $\left[\frac{1}{m}, \frac{1}{m-1}\right)$. Going from left to right, let $(G_{m,n}(\omega),\ D_{m,n}(\omega))$ be the nth such interval if it exists; otherwise, put $G_{m,n}(\omega) = D_{m,n}(\omega) = +\infty$ and note that the interval becomes empty. Finally, to lighten the notation, use a bijection $(m, n) \mapsto i$ from $\mathbb{N}^* \times \mathbb{N}^*$ onto $\mathbb{N}$ to re-label these intervals as $(G_i(\omega),\ D_i(\omega))$. Thus,

5.2 $$\mathbb{R}_+\backslash C_\omega = \bigcup_{i\in\mathbb{N}} (G_i(\omega), D_i(\omega)), \qquad \omega \in \Omega.$$

Clearly, each D_i is a stopping time. Stability and recurrence properties of W imply that each D_i is almost surely finite. Incidentally, each G_i is a random variable but not a stopping time; see Exercise 2.27 for the reasoning. The following shows the Cantor set like features of the zero-set C.

5.3 THEOREM. *For almost every ω, the set C_ω is perfect and unbounded, its interior is empty, its Lebesgue measure is zero, and it has the power of the continuum.*

Proof. We have already seen that C_ω is closed. It is unbounded for almost every ω in view of Corollary 2.4. Its Lebesgue measure is zero for almost every ω, since

$$\mathbb{E}\ \text{Leb}\ C = \mathbb{E} \int_{\mathbb{R}_+} dt\ 1_{\{0\}} \circ W_t = \int_{\mathbb{R}_+} dt\ \mathbb{P}\{W_t = 0\} = 0.$$

This implies that the interior of C_ω is empty for almost every ω, because no set of zero Lebesgue measure can contain an open interval. To complete the proof, there remains to show that, for almost every ω, the set C_ω has no isolated points; then, the closed set C_ω is perfect and has the power of the continuum necessarily.

We start by recalling that $T_0 = 0$ almost surely. Thus, as mentioned in Corollary 2.3, there is an almost sure set Ω_{00} such that, for every ω in it, there is a strictly decreasing sequence (t_k) in C_ω with limit 0, that is, the point 0 of C_ω is a limit point of C_ω for every ω in Ω_{00}.

Similarly, for each i in $\mathbb{N}$, the stopping time D_i is almost surely finite, and the strong Markov property yields that

$$T_0 \circ D_i = 0 \qquad \text{almost surely.}$$

Thus, there is an almost sure event Ω_i such that $D_i(\omega)$ is a limit point of C_ω for every ω in Ω_i. Consider, finally, the intersection Ω' of the events $\Omega_{00},\ \Omega_0,\ \Omega_1,\ \ldots$. For ω in it, neither 0 nor any $D_i(\omega)$ is isolated. In view of 5.2, then, C_ω is perfect for every ω in the almost sure event Ω'. □

5.4 REMARKS. a) It will be convenient to introduce here the almost sure event $\Omega^* = \Omega' \cap \Omega''$, where Ω' is as in the proof above, and where Ω'' is the set of ω for which the claims of the preceding theorem hold in addition to the regularity properties of the path $W(\omega)$.

b) We shall see in Corollary 5.11 below that there is a strictly increasing function $a \mapsto S_a(\omega)$ from $\mathbb{R}_+$ into $\mathbb{R}_+$ such that $S_a(\omega)$ belongs to C_ω for every a. This shows, directly, that C_ω has at least "as many points" as $\mathbb{R}_+$.

Local time at zero

Imagine a clock whose mechanism is so rigged that the clock advances when and only when the Wiener particle is at the origin. We shall show that such a clock exists; it will be called the *local time at zero.*

First, some generalities. Let $c:\ \mathbb{R}_+ \mapsto \mathbb{R}_+$ be increasing and continuous with $c(0) = 0$. Think of it as a clock: when the standard time is t, the clock shows $c(t)$. The clock may remain stationary during some periods of time, that is, the function is not forced to be strictly increasing. The *set of times of increase for c* is defined to be

5.5 $$\text{Incr}\ c = \{t \in \mathbb{R}_+ : c(t-\varepsilon) < c(t+\varepsilon) \quad \text{for every } \varepsilon > 0\},$$

where we use the convention that $c(t-\varepsilon)=0$ for $t<\varepsilon$. Corresponding to c there is a unique measure on $\mathbb{R}_+$ whose "distribution" function is c. The set Incr c is also called the *support* of this measure, namely, the smallest closed set whose complement has measure zero. We shall show next the existence of a random measure on $\mathbb{R}_+$ whose support is the zero-set C defined by 5.1. This is interesting especially since C has zero as its Lebesgue measure.

Consider Figure 13 on page 402, and concentrate on the relationship of M to the Wiener process V there; recall that $|V|=M-W$. For every ω, the path $M(\omega)$ is increasing and continuous; and it increases at a time t if only if $V_t(\omega)=0$, more precisely,

5.6 $$\text{Incr } M(\omega)=\{t\in\mathbb{R}_+ : V_t(\omega)=0\}\,.$$

Moreover, it follows from Theorem 4.4 that the time-set on the right side determines the path $M(\omega)$.

Since W is a Wiener process just as V, there must be a process L that is related to W just as M is to V. We state this conclusion next; there is nothing new to prove. The process $L=(L_t)_{t\in\mathbb{R}_+}$ is called the *local time* of W at zero.

5.7 THEOREM. *There exists an increasing continuous process L that has the same law as M and is such that*

$$\text{Incr } L(\omega)=C_\omega=\{t\in R_+ : W_t(\omega)=0\}\,,\qquad \omega\in\Omega.$$

Inverse of the local time is a stable Lévy process

Heuristically, the local time process L is a random clock that advances when and only when W is at the point 0. When the standard time is t, the local time at 0 is L_t; conversely,

5.8 $$S_a=\inf\{t\in\mathbb{R}_+ : L_t>a\}$$

is the standard time when the local time is just about to pass a.

5.9 THEOREM. *The process $S=(S_a)_{a\in\mathbb{R}_+}$ has the same law as the hitting time process $T=(T_a)_{a\in\mathbb{R}_+}$. It is a strictly increasing pure-jump Lévy process; it is stable with index $1/2$; its Lévy measure is*

$$\lambda(ds)=ds\frac{1}{\sqrt{2\pi s^3}},\qquad s>0,$$

Proof. By comparing 5.8 and 3.2, we note that S bears the same relation to L as T does to M. By the last theorem, L and M have the same law. Hence, S and T have the same law. The statement about S as a Lévy process is the same as Theorem 3.6 about T. □

In terms of S, the local time L is defined as the functional inverse of S:

5.10 $$L_t = \inf\{a : S_a > t\}, \qquad t \in \mathbb{R}_+;$$

This is immediate from 5.8. The following is to clarify some further relationships.

5.11 COROLLARY. *For almost every ω, with $G_i(\omega)$ as in 5.2,*

$$\{S_a(\omega) : a \in \mathbb{R}_+\} = C_\omega \setminus \{G_i(\omega) : i \in \mathbb{N}\}.$$

Proof. Take ω such that $L(\omega)$ is continuous and $S(\omega)$ strictly increasing. Fix a in $\mathbb{R}_+$ and let $S_a(\omega) = t$. Then, 5.8 and the continuity of $L(\omega)$ imply that

5.12 $$S_a(\omega) = t \Leftrightarrow L_t(\omega) = a, \quad L_{t+\varepsilon}(\omega) > a \quad \text{for every } \varepsilon > 0.$$

Next, note that the set of t for which the right side holds for some a is exactly the set $(\text{Incr } L(\omega))\setminus\hat{G}_\omega$, where $\hat{G}_\omega$ is the countable set consisting of the left-end-points of the intervals contiguous to Incr $L(\omega)$. The proof is complete, since $C_\omega = \text{Incr } L(\omega)$ by Theorem 5.7 and, thus, $\hat{G}_\omega = \{G_i(\omega) = i \in \mathbb{N}\}$ by 5.2. □

Local times elsewhere

Fix a point x in $\mathbb{R}$. Consider the hitting time T_x defined by 2.1 or 2.14. It is almost rurely finite; the Wiener particle is at x at that time; and the point x becomes the point 0 of the new Wiener process $W \circ \theta_{T_x}$. With L as defined earlier, $L \circ \theta_{T_x}$ is the local time at 0 for $W \circ \theta_{T_x}$; using it, we introduce the following definition for every outcome ω and time t:

5.13 $$L_t^x(\omega) = \begin{cases} 0 & \text{if} \quad t < T_x(\omega), \\ L_{t-s}(\theta_s\omega) & \text{if} \quad t \geq s = T_x(\omega). \end{cases}$$

It is immediate from Theorem 5.7 that

5.14 $$\text{Incr } L^x(\omega) = \{t \in \mathbb{R}_+ : W_t(\omega) = x\}.$$

Thus, the process $L^x = ({L^x}_t)_{t\in\mathbb{R}_+}$ is called the local time at x for W. Note that $L^0 = L$.

For $x \neq 0$, the path $L^x(\omega)$ stays at 0 during $[0,\ T_x]$ and, then, starts increasing just as L did at 0. All computations regarding L^x can be reduced to computations about L, but with special consideration for the delay at the start; see Exercise 5.19 for an example.

Master theorem on equivalence

The essential argument underlying the results of this section is that L bears the same relationship to W as M does to V. We put this observation next and supplement it by recalling that $|V| = M - W$. This is a summary of the results above; there is nothing new to prove.

5.15 THEOREM. *The three-dimensional process $(W,\ L,\ S)$ has the same law as $(V,\ M,\ T)$. Further, $(|W|,\ L,\ S)$ has the same law as $(M-W,\ M,\ T)$, and $L-|W|$ has the same law as W.*

Exercises

5.16 *Minimum of W.* Define $m_t(\omega) = \min_{0\le s\le t}\ W_s(\omega)$. Obviously, the process $(-m_t)$ has the same law of (M_t). Show that $|W|$ has the same law as $W-m$.

5.17 *Local time measure.* For each ω, let $A \mapsto L(\omega, A)$ be the unique measure on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ whose distribution function is the increasing continuous function $t \mapsto L_t(\omega)$. Show that the support of that measure is exactly the set C_ω of zeros of $W(\omega)$–the support of a measure μ on $\mathbb{R}_+$ is the smallest closed subset of $\mathbb{R}_+$ whose complement has μ-measure 0. Obviously, $L(\omega,\cdot)$ is singular with respect to the Lebesgue measure, and

$$L(\omega,[0,t]) = L(\omega, C_\omega \cap [0,t]) = L_t(\omega).$$

5.18 *Computing the local time.* This is to relate L to the zero-set C. For every ω and every $\varepsilon > 0$, let $\hat{N}_t(\omega,\varepsilon)$ be the number of intervals that are contiguous to C_ω, are contained in $[0,t]$, and whose lengths exceed ε. Show that, for almost every ω,

$$\lim_{\varepsilon\downarrow 0}\ \sqrt{2\pi\varepsilon}\ \hat{N}_t(\omega,\varepsilon) = 2\ L_t(\omega).$$

5.19 *Local time at x.* Note that L^x and L^{-x} have the same law for every x in $\mathbb{R}$. Fix $x > 0$. Compute $\mathbb{P}\{L_t^x = 0\}$. Show that, for $a > 0$,

$$\mathbb{P}\{L_t^x \in da\} = \int_{[0,t]} \mathbb{P}\{T_x \in ds\}\,\mathbb{P}\{L_{t-s} \in da\} = da\frac{2e^{-(x+a)^2/2t}}{\sqrt{2\pi t}}.$$

5.20 *Inverse of the local time at x.* Fix x in $\mathbb{R}$. Define S_a^x from L^x as S_a is defined from L in 5.8. Show that the process $(S_a^x)_{a\in\mathbb{R}_+}$ has the same probability law as $(T_x + \hat{S}_a)_{a\in\mathbb{R}_+}$, where $(\hat{S}_a)$ is independent of T_x and has the same law as the stable Lévy process (S_a) described in Theorem 5.9.

5.21 *Occupation times.* For x in $\mathbb{R}$ and t in $\mathbb{R}_+$, define

$$A_t(\omega,x) = \int_0^t\ ds\ 1_{(-\infty,x]}\circ W_s(\omega),$$

the amount of time spent in $(-\infty,\ x]$ by $W(\omega)$ during $[0,\ t]$. Show that $x \mapsto A_t(\omega,\ x)$ is equal to 0 on $(-\infty,\ m_t(\omega)]$, and to t on $[M_t(\omega), +\infty)$, and is continuous and strictly increasing on $[m_t(\omega),\ M_t(\omega)]$; what are $m_t(\omega)$ and $M_t(\omega)$?

5.22 *Local times as derivatives.* It can be shown that, for almost every ω, $x \mapsto A_t(\omega,\ x)$ is differentiable, and its derivative at the point x is $L_t^x(\omega)$, that is,

$$L_t^x(\omega) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t ds\ 1_{(x-\varepsilon, x+\varepsilon)} \circ W_s(\omega).$$

5.23 *Occupation measure.* This is the name for the measure on $\mathbb{R}$ whose distribution function is $x \mapsto A_t(x)$. Letting it be denoted by K_t, we see from 5.21 and 5.22 that

$$K_t(\omega, B) = \int_0^t ds\ 1_B \circ W_s(\omega) = \int_B dx\ L_t^x(\omega), \quad B \in \mathcal{B}_{\mathbb{R}},$$

or, for f positive Borel on $\mathbb{R}$,

$$K_t f(\omega) = \int_0^t ds\ f \circ W_s(\omega) = \int_{\mathbb{R}} dx\ f(x)\ L_t^x(\omega).$$

5.24 *Continuity of local times.* It is known that, for almost every ω, the mapping

$$(x, t) \mapsto L_t^x(\omega)$$

from $\mathbb{R} \times \mathbb{R}_+$ into $\mathbb{R}_+$ is continuous.

6 Excursions

We continue with the setup and notations of the previous sections: W is the Wiener process under consideration, C is its set of zeros, L is its local time process at 0, and S is the inverse local time. Recall the almost sure event introduced in Remark 5.4a; we take it to be the new Ω in order to avoid boring repetitions of "almost every." We are interested in the excursions of W outside the point 0, that is, basically, in the path segments over the intervals contiguous to C.

Excursion space

The path segments in question are continuous functions that start from 0, stay away from 0 for some strictly positive time, and return to 0 some finite time later. It is convenient to let each such function remain at zero forever after the return to 0. The following is the space of such functions.

We define E to be the collection of all continuous functions $x : \mathbb{R}_+ \mapsto \mathbb{R}$ such that

6.1 $$\zeta(x) = \inf\{t > 0 : x(t) = 0\}$$

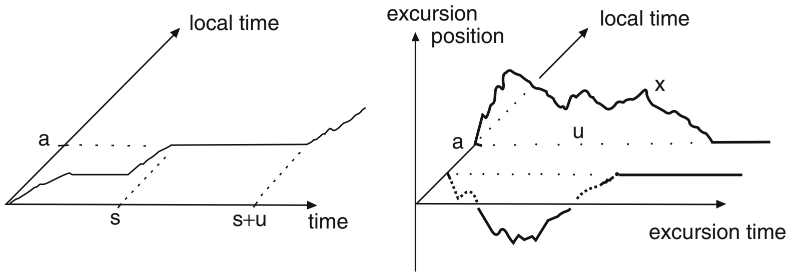


Figure 14: At the local time a there is an excursion x of duration u. There are infinitely many excursions, one for each flat stretch of the local time clock.

is a strictly positive real number, and x vanishes outside $(0, \zeta(x))$. Each element x of E is called an *excursion*; $\zeta(x)$ is its *duration*; note that x is either a positive function or negative. We let $\mathcal{E}$ be the Borel σ-algebra on E corresponding to the topology of uniform convergence on compacts. Then, $(E, \mathcal{E})$ is called the *excursion space*.

Excursions of W

Fix an outcome ω. Let a be a point on the local time axis (see Figure 14 below) at which $S(\omega)$ has a jump, say, from $s = S_{a-}(\omega)$ to $s+u = S_a(\omega)$, with $u > 0$. During the interval $[s,\ s+u]$ the local time $L(\omega)$ stays constant at the value a, and the Wiener path $W(\omega)$ has an excursion x defined formally by

6.2 $$x(t) = \begin{cases} W_{s+t}(\omega) & \text{if} \quad 0 \le t \le u \\ 0 & \text{if} \quad t > u. \end{cases}$$

This x is called the excursion of $W(\omega)$ at the local time a. It is an element of E, and its duration is

6.3 $$\zeta(x) = u = S_a(\omega) - S_{a-}(\omega),$$

which is strictly positive and finite by the way a is chosen. Each excursion corresponds to a local time at which $S(\omega)$ jumps.

Poisson random measure for excursions

The next theorem is fundamental. The measure ν on $(E, \mathcal{E})$ describing the mean here is called the *Itô measure* of excursions. We shall specify it later.

6.4 THEOREM. *For every ω, let $N(\omega, \cdot)$ be the counting measure on $(\mathbb{R}_+ \times E, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{E})$ whose atoms are the pairs (a, x), where x is an excursion of $W(\omega)$, and a the corresponding local time. Then, N is a Poisson random measure whose mean has the form* Leb $\times\, \nu$, *where ν is a σ-finite measure on the excursion space $(E,\ \mathcal{E})$.*

Proof. Fix $\varepsilon > 0$. Let $A_1,\ A_2,\ \dots$ be the successive points of jump for $a \mapsto S_a$ with jump sizes exceeding ε. By Theorem 5.9, these A_i form a Poisson random measure on $\mathbb{R}_+$ with mean c_ε Leb, where $c_\varepsilon = \lambda(\varepsilon, \infty) = 2/\sqrt{2\pi\varepsilon}$.

Corresponding to the local time A_i, let X_i be the excursion, and $D_i = S_{A_i}$ the right-end point of the contiguous interval over which the local time is A_i. Each X_i is a random variable taking values in $(E, \mathcal{E})$. Each D_i is a finite stopping time with $W_{D_i} = 0$. Since $0 < A_1 < A_2 < \dots$, we have $0 < D_1 < D_2 < \dots$, and $A_1,\ X_1,\ \dots,\ A_i,\ X_i$ belong to the past $\mathcal{F}_{D_i}$. By the strong Markov property at D_i, then, the pair $(A_{i+1} - A_i,\ X_{i+1})$ is independent of $\mathcal{F}_{D_i}$ and, therefore, of $\{A_1, X_1,\ \dots,\ A_i,\ X_i\}$, and has the same distribution as $(A_1,\ X_1)$. Noting further that A_1 and X_1 are independent, we conclude the following: (A_i) forms a Poisson random measure on $\mathbb{R}_+$ with mean c_ε Leb; (X_i) is independent of it and is an independency of variables with some distribution μ_ε on $(E, \mathcal{E})$ in common. It follows from Corollary VI.3.5 that the pairs $(A_i,\ X_i)$, $i \geq 1$, form a Poisson random measure N_ε on $\mathbb{R}_+ \times E$ whose mean measure is Leb $\times\, \nu_\varepsilon$, where $\nu_\varepsilon = c_\varepsilon\, \mu_\varepsilon$ is a finite measure on $(E, \mathcal{E})$.

Observe that, for every ω, the atoms $(A_i(\omega),\ X_i(\omega))$ are those atoms (a, x) of $N(\omega, \cdot)$ with $\zeta(x) > \varepsilon$. Thus, the Poisson random measure N_ε is the trace of N on $\mathbb{R}_+ \times E_\varepsilon$, where $E_\varepsilon = \{x \in E : \zeta(x) > \varepsilon\}$. Letting $\varepsilon \to 0$, we conclude that N is a Poisson random measure on $R_+ \times E$ whose mean measure is Leb $\times\ \nu$, where ν is the measure defined by

$$\nu f = \lim_{\varepsilon \to 0} \nu_\varepsilon f, \qquad f \in \mathcal{E}_+.$$

Since $\nu(E_\varepsilon) = \nu_\varepsilon(E) = c_\varepsilon < \infty$ for every $\varepsilon > 0$, the measure ν is σ-finite. □

Excursions determine W

We have constructed the Poisson random measure N above, ω by ω, from the Wiener process W. This can be reversed: N determines W.

Recall from 6.3 that the duration $\zeta(x)$ of an excursion x is the jump size for $S(\omega)$ at the local time corresponding to that excursion. Thus, for every ω,

6.5 $$S_a(\omega) = \int_{[0,a]\times E} N\,(\omega; db, dx)\,\zeta(x), \qquad a \in \mathbb{R}_+,$$

and $L(\omega)$ is the functional inverse of $S(\omega)$; and

6.6 $$W_t(\omega) = \int_{[0, L_t(\omega)]\times E} N\,(\omega; da, dx)\, x\,(t - S_{a-}(\omega))\,, \qquad t \in \mathbb{R}_+.$$

In fact, the last integral is a countable sum with at most one non-zero term, namely, the term corresponding to $a = L_t(\omega)$ if $S_{a-}(\omega) < S_a(\omega)$.

Extents of excursions

In preparation for characterizing Itô's measure ν on the excursion space $(E, \mathcal{E})$, we describe next the law it imparts on the extents of excursions.

For an excursion x in E, we define the *extent* of x to be the point touched by x that is at maximum distance from 0, that is,

6.7 $$m(x) = \begin{cases} \max_{t\in\mathbb{R}_+} x(t) & \text{if} \quad x \text{ is positive,} \\ \min_{t\in\mathbb{R}_+} x(t) & \text{if} \quad x \text{ is negative;} \end{cases}$$

recall that x is either positive or negative. The next theorem shows that the local times and extents of excursions form a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with an explicit mean measure.

6.8 THEOREM. *Let h be the mapping $(a, x) \mapsto (a,\ m(x))$ from $\mathbb{R}_+ \times E$ into $\mathbb{R}_+ \times \mathbb{R}$. Then, $\hat{N} = N \circ h^{-1}$ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ whose mean has the form* Leb $\times\, \hat{\nu}$, *where*

$$\hat{\nu}(db) = db\ \frac{1}{2b^2}, \qquad b \in \mathbb{R}.$$

REMARK. It is curious that $\hat{\nu}$ is the Lévy measure of a Cauchy process, namely, $(1/2\pi Y_t)$ where Y is standard Cauchy process; see Example VII.2.14.

Proof. a) Since $m : E \mapsto \mathbb{R}$ is continuous, the mapping h is measurable with respect to the Borel σ-algebras on $\mathbb{R}_+ \times E$ and $\mathbb{R}_+ \times \mathbb{R}$. Since N is Poisson on $\mathbb{R}_+ \times E$ with mean Leb $\times\, \nu$, it follows that $\hat{N}$ is Poisson on $\mathbb{R}_+ \times \mathbb{R}$ with mean Leb $\times\, \hat{\nu}$, where $\hat{\nu} = \nu \circ m^{-1}$. By the symmetry of W, the measure $\hat{\nu}$ on $\mathbb{R}$ must be symmetric. Thus, for every $b > 0$,

6.9 $$\hat{\nu}(b, \infty) = \hat{\nu}(-\infty, -b) = \frac{1}{2}\left[\hat{\nu}(b, \infty) + \hat{\nu}(-\infty, -b)\right] = \frac{1}{2}\nu(E_b),$$

where

$$E_b = \{x \in E : |m(x)| > b\}\,.$$

To complete the proof, we shall show that $\nu(E_b) = 1/b$.

b) Fix $b > 0$ and define

$$\tau = \inf\{t : |W_t| > b\}\,;$$

recall that $\mathbb{E}\tau = b^2$. Consider L_τ, the local time at the standard time τ. Note that it is also the local time corresponding to the first excursion belonging to E_b. Thus, for every ω,

$$L_{\tau(\omega)}(\omega) > a \Leftrightarrow N(\omega, [0, a] \times E_b) = 0.$$

Since N is Poisson with mean Leb $\times\, \nu$, then,

6.10 $$\mathbb{P}\{L_\tau > a\} = \exp_\ a\ \nu(E_b), \qquad a \in \mathbb{R}_+.$$

c) For the same $b > 0$, define

$$\sigma = \inf\{t : M_t - W_t > b\}\,.$$

Since $(|W|,\ L)$ has the same law as $(M-W,\, M)$ by Theorem 5.15, we deduce that $(\tau,\ L_\tau)$ has the same distribution as $(\sigma,\ M_\sigma)$. Hence,

6.11 $$\mathbb{E}\,\sigma = \mathbb{E}\,\tau = b^2,\ \mathbb{E}\,M_\sigma = \mathbb{E}\ L_\tau = 1/\nu(E_b),$$

the last equality being a consequence of 6.10. Since $\sigma < \infty$ almost surely, $M_\sigma - W_\sigma = b$ by the definition of σ and the continuity of $M - W$. Hence, to complete the proof via 6.11 and 6.9, there remains to show that

6.12 $$\mathbb{E}\,W_\sigma = 0.$$

d) Consider the martingale $X = \left(W_t^2 - t\right)_{t\in\mathbb{R}_+}$. For each t, it is Doob on $[0,\,t]$ by V.5.6, and thus, $\mathbb{E}X_{\sigma\wedge t} = 0$. Hence,

$$\mathbb{E}\,W_{\sigma\wedge t}^2 = \mathbb{E}\ (\sigma\wedge t) \le \mathbb{E}\ \sigma = b^2, \qquad t\in\mathbb{R}_+,$$

which shows that the martingale $(W_{\sigma\wedge t})_{t\in\mathbb{R}_+}$ is L^2-bounded and, thus, uniformly integrable (see Remark II.3.13e). By Theorem V.5.14, this is equivalent to saying that the martingale W is Doob on $[0,\ \sigma]$. Hence, 6.12. □

Itô measure on excursions

Recall the Poisson random measure N of excursions; see Theorem 6.4. Its mean measure on $\mathbb{R}_+ \times E$ is the product measure Leb $\times\, \nu$, where ν is a σ-finite measure on $(E, \mathcal{E})$. Our aim is to state a characterization for ν, the Itô measure.

Let $(A_i,\ X_i),\ i \in \mathbb{N}$, be an enumeration of the atoms of $\mathbb{N}$, that is, the pairs $(A_i,\ X_i)$ are random variables taking values in $\mathbb{R}_+ \times E$, and they form N. Then, the pairs $(A_i,\ m\circ X_i)$ are the atoms of the Poisson random measure $\hat{N}$ described in Theorem 6.8. Finally, the triplets $(A_i,\ m\circ X_i,\ X_i)$ must form a Poisson random measure $\tilde{N}$, namely, $\tilde{N} = N\circ h^{-1}$ where $h(a,\ x) = (a,\ m(x),\ x)$. The following is immediate from Theorems 6.4 and 6.8; no proof is needed.

6.13 Proposition. *The mean μ of the Poisson random measure $\tilde{N}$ is given by*

6.14 $$\mu(da, db, dx) = da\ db\ \frac{1}{2\,b^2}\ Q(b, dx), \qquad a\in\mathbb{R}_+, b\in\mathbb{R}, x\in E,$$

where Q is the transition probability kernel from $(\mathbb{R},\ \mathcal{B}_\mathbb{R})$ into $(E, \mathcal{E})$ defined by

6.15 $$Q(b, D) = \mathbb{P}\left\{X_i \in D | m\circ X_i = b\right\}, \qquad b\in\mathbb{R},\ D\in\mathcal{E}.$$

6.16 Corollary. *Itô measure ν for excursions is given by*

$$\nu(D) = \int_\mathbb{R} db\ \frac{1}{2\,b^2}\ Q(b, D), \qquad D\in\mathcal{E}.$$

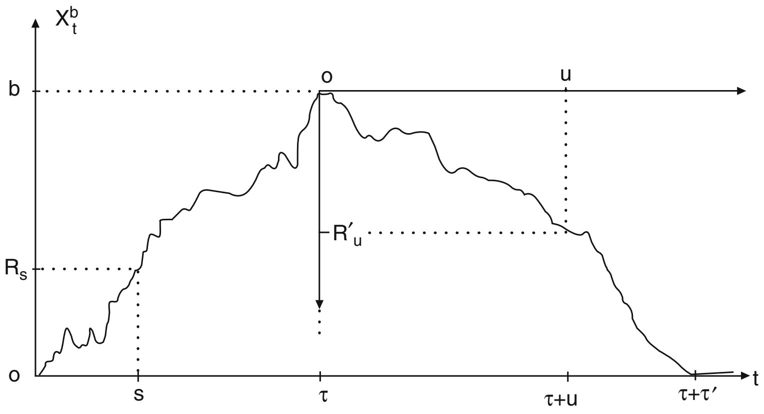


Figure 15: Excursion X^b with a given extent $b > 0$. Run a bessel process R upward until it hits b; then run a new Bessel process R' downward from b until it hits 0.

Proof is immediate from the form 6.14 for the mean of $\tilde{N}$, since $N = \tilde{N} \circ h^{-1}$ with $h(a, b, x) = (a, x)$, which implies that $\text{Leb} \times \nu = \mu \circ h^{-1}$. □

The preceding corollary reduces the task of characterizing the Itô measure ν to that of characterizing the probability measure $Q(b, \cdot)$ for each b, namely, the probability law of an excursion whose extent is given to be b.

It is obvious that $Q(b,\ D) = Q(-b,\ -D)$ for $b < 0$, with $-D = \{-x :\ x \in D\}$; this is by the symmetry of W. It is also obvious that, if $b > 0$, then $Q(b, \cdot)$ must put all its mass on the set of positive excursions. Thus, the following characterization specifies Q completely, and via the last corollary, Itô measure ν. See Figure 15 as well. This theorem of D. Williams's is put here without proof; see the notes for this chapter.

6.17 THEOREM. *Let R and R' be independent Bessel processes with index 3. Let τ_b be the hitting time of the level $b > 0$ by R, and τ'_b the same for R'. Define, for ω in Ω and t in $\mathbb{R}_+$,*

6.18 $$X_t^b(\omega) = \begin{cases} R_t(\omega) & \text{if} \quad 0 \le t \le \tau_b(\omega) \\ b - R'_{t-\tau_b(\omega)}(\omega) & \text{if} \quad \tau_b(\omega) \le t \le \tau_b(\omega) + \tau'_b(\omega) \\ 0 & \text{if} \quad t > \tau_b(\omega) + \tau'_b(\omega). \end{cases}$$

Then, $Q(b, \cdot)$ is the probability law of the process X^b.

The preceding theorem together with Corollary 6.16 characterizes the Itô measure in terms of well-understood operations. For Bessel processes see Example 1.22; recall that R here is the radial part of a three-dimensional Wiener process. See also Theorem 4.19, which shows that R has the same law as $2M - W$.

Local times of some hits

This is to expand on the observation, within the proof of Theorem 6.8, that the local time L_τ at the time τ of exit from $(-b,\ b)$ has the exponential distribution with mean b. Recall that T_a is the time W hits (a,∞) if $a \ge 0$, and is the time of hitting $(-\infty,\ a)$ if $a \le 0$.

6.19 Proposition. *Let $a,\ b > 0$. Then, L_{T_a} and $L_{T_{-b}}$ are independent and exponentially distributed with means $2a$ and $2b$ respectively. Moreover, $L_{T_a \wedge T_{-b}}$ is equal to $L_{T_a} \wedge L_{T_{-b}}$ and has the exponential distribution with mean $2ab/(a+b)$.*

Proof. In terms of the Poisson random measure N of excursions, we have

6.20 $$\left\{L_{T_a} > u,\ L_{T_{-b}} > v\right\} = \{N([0,u] \times A) = 0\} \cap \{N([0,v] \times B) = 0\}$$

where

$$A = \{x \in E : m(x) > a\}, \qquad B = \{x \in E : m(x) < -b\}.$$

Since A and B are disjoint, the right side of 6.20 is the intersection of two independent events. Hence, by 6.4,

$$\mathbb{P}\left\{L_{T_a} > u, L_{T_{-b}} > v\right\} = e^{-u\nu(A)} e^{-v\nu(B)}, \quad u, v \in \mathbb{R}_+,$$

where, by Theorem 6.8,

$$\nu(A) = 1/2a, \qquad \nu(B) = 1/2b.$$

This proves the first statement. The second is immediate from it and the computation $\nu(A) + \nu(B) = (a+b)/2ab$. □

The arcsine law

As another illustration of the uses of excursion theory, we prove next *the* arcsine law, the most celebrated of the arcsine laws. It specifies the distribution of

6.21 $$A_t = \int_{[0,t]} ds\ 1_{\mathbb{R}_+} \circ W_s, \qquad t \in \mathbb{R}_+,$$

and is the main ingredient in computations about occupation times and Brownian quantiles; see Exercises 6.41–6.47.

6.22 Theorem. *The distribution of A_t is the same as that of tA, where A has the arcsine distribution as in 2.15.*

6.23 Remark. In view of the (simpler to obtain) arcsine law given in Proposition 2.18, we see that G_t and A_t have the same distribution. See Exercise 6.40 for the underlying reasons.

Proof. Consider the standard time S_a corresponding to the local time a. It is obtained via 6.5 from the Poisson random measure N of excursions. Then, (S_a) is a pure-jump Lévy process whose Lévy measure λ is given by (see 6.5 and 5.9)

6.24 $$\lambda f = \int_0^\infty ds \; \frac{1}{\sqrt{2\pi s^3}} f(s) = \int_E \nu(dx)\; f(\zeta(x)),$$

where ν is the Itô measure of excursions. We now decompose S as

$$S = S^+ + S^-,$$

where S_a^+ is the time spent on positive excursions during $[0, S_a]$, and S_a^- is that on negative excursions:

6.25 $$S_a^+ = \int_{[0,a]\times E_+} N(db, dx)\; \zeta(x), \qquad \text{where } E_+ = \{x \in E : x \geq 0\},$$

and S^- is defined similarly but with $E_- = \{x \in E : x \leq 0\}$.

Since E_+ and E_- are disjoint, and since N is Poisson, the processes S^+ and S^- are independent. Comparing 6.25 with 6.5, we see that S^+ and, by symmetry, S^- are pure-jump Lévy processes with the same Lévy measure, namely, $\frac{1}{2}\lambda$. We conclude that S_a^+ and S_a^- are independent and have the same distribution as $S_{a/2}$, that is,

$$\mathbb{P}\left\{S_a^+ \in du, S_a^- \in dv\right\} = du\; dv\; \frac{a\; e^{-a^2/8u}}{2\sqrt{2\pi u^3}} \cdot \frac{a\; e^{-a^2/8v}}{2\sqrt{2\pi v^3}}.$$

Hence, for positive Borel functions f on $\mathbb{R}_+ \times \mathbb{R}_+$, an easy computation yields

6.26 $$\mathbb{E} \int_{\mathbb{R}_+} da\; f\left(S_a^+, S_a^-\right) = \int_{\mathbb{R}_+} du \int_{\mathbb{R}_+} dv \frac{1}{\sqrt{2\pi(u+v)^3}}\; f(u,v).$$

b) It follows from the scaling property of W that A_t has the same distribution as tA_1. Thus, we concentrate on the distribution of A_1.

Fix $a > 0$; for almost every ω, the counting measure $N(\omega, \cdot)$ has exactly one atom (a, x) such that $S_{a-}(\omega) \leq 1 < S_a(\omega) = S_{a-}(\omega) + \zeta(x)$, and, then,

$$A_1(\omega) = \begin{cases} S_{a-}^+(\omega) & \text{if } \; x \in E_-, \\ 1 - S_{a-}^-(\omega) & \text{if } \; x \in E_+. \end{cases}$$

In other words, for Borel $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$,

6.27 $$f(A_1) = \int_{\mathbb{R}_+ \times E} N(da, dx)\; g(S_{a-}^+, S_{a-}^-, x)$$

where

$$g(u, v, x) = 1_{[0,1]}(u+v)\; 1_{(1,\infty)}(u+v+\zeta(x))[f(u)1_{E_+}(x) + f(1-v)1_{E_-}(x)].$$

Applying Theorem VI.6.2 to the Poisson integral in 6.27, recalling that the mean of N is Leb $\times\,\nu$, we see that

$$\mathbb{E}f(A_1) = \mathbb{E}\int_{\mathbb{R}_+} da \int_E \nu(dx) \quad g\left(S_a^+, S_a^-, x\right)$$

6.28
$$= \int_E \nu(dx) \int_{\mathbb{R}_+} du \int_{\mathbb{R}_+} dv \frac{1}{\sqrt{2\pi(u+v)^3}}\; g(u,v,x),$$

where we used 6.26 at the last step. In view of 6.24,

$$\int_E \nu(dx) g(u,v,x) = 1_{[0,1]}(u+v)\int_0^\infty ds \frac{1}{\sqrt{2\pi s^3}} 1_{(1,\infty)}(u+v+s) \times \left[\tfrac{1}{2}f(u) + \tfrac{1}{2}f(1-v)\right] = 1_{[0,1]}(u+v)\frac{1}{\sqrt{2\pi(1-u-v)}}\left[f(u)+f(1-v)\right].$$

Putting this into 6.28 we obtain, with $s = u+v$ and $r = u/s$,

$$\mathbb{E}f(A_1) = \int_0^1 ds\; \frac{1}{\pi\sqrt{s(1-s)}} \int_0^1 dr \left[\tfrac{1}{2}f(sr) + \tfrac{1}{2}\, f(1-s+sr)\right]$$

6.29
$$= \mathbb{E}\left[\tfrac{1}{2}f(AU) + \tfrac{1}{2}\, f(1-A+AU)\right],$$

where A and U are independent, A has the arcsine distribution as in 2.15, and U is uniform on (0, 1).

It is easy to show that, then, AU has the beta distribution with parameter ($\tfrac{1}{2}$, $\tfrac{3}{2}$), and so does $A - AU = A(1-U)$ since $1-U$ is also uniform on (0, 1); see Exercise 6.39. Hence, 6.29 yields

$$\mathbb{E}\; f(A_1) = \textstyle\int_0^1 dv\; \frac{2}{\pi} v^{-1/2}(1-v)^{1/2} \left[\tfrac{1}{2}\; f(v) + \tfrac{1}{2}\; f(1-v)\right] = \textstyle\int_0^1 du \frac{1}{\pi\sqrt{u(1-u)}}\; f(u).$$

This proves that A_1 has the arcsine distribution as does A in 2.15, which completes the proof since A_t and tA_1 have the same distribution. □

Exercises

Notation: W, M, T, L, S, N, V retain the meanings they had within the present section and earlier. Below, for random variables X and Y (or processes X and Y), we write $X \approx Y$ to mean that X and Y have the same distribution. Throughout, A will denote a random variable having the arcsine distribution as in 2.15.

6.30 *Skew Brownian motion.* Recall the Itô measure ν regulating the excursions, and the sets E_+ and E_- of positive and negative excursions. Let ν_+ be the trace of 2ν on E_+, and ν_- the trace of 2ν on E_-. Then,

$$\nu(D) = \nu(D \cap E_+) + \nu(D \cap E_-) = \frac{1}{2}\nu_+(D) + \frac{1}{2}\nu_-(D), \qquad D \in \mathcal{E}.$$

This is a precise expression of the heuristic that each excursion is positive with probability $1/2$ and negative with $1/2$. Define, for $0 < p < 1$ and $q = 1 - p$, a new measure on $(E, \mathcal{E})$. Let

$$\nu^* = p\ \nu_+ + q\ \nu_-,$$

and let N^* be the Poisson random measure on $\mathbb{R}_+ \times E$ with mean measure $\text{Leb} \times \nu^*$. Define W^* from N^* as W is defined from N through 6.5 and 6.6. The resulting process W^* is called skew Brownian motion; it is a Markov process. It is *not* symmetric, its increments are not independent. Find the distribution of W_t^*. Compute its transition function (P_t).

6.31 *Random time changes.* Many interesting Markov processes are obtained from Wiener processes by random time changes. Here is the general setup. Let $H = (H_t)$ be a random clock; assume that $t \mapsto H_t(\omega)$ is increasing and continuous, starting from $H_0(\omega) = 0$. We think of H_t as the clock time when the standard time is t. Then,

$$\tau_u = \inf\{t : H_t > u\}$$

is the standard time when the clock reads u, and

$$X_u = W_{\tau_u}$$

is the position of the Wiener particle at that time. The simplest case is when $t \mapsto H_t$ is deterministic, strictly increasing, and continuous, then X has (possibly non-stationary) independent increments. Following are some special cases.

6.32 *Reflected Brownian motion.* In 6.31, Suppose that

$$H_t = \int_0^t ds\ 1_{\mathbb{R}_+} \circ W_s.$$

show that X is a reflected Brownian motion, that is, $X \approx |W|$.
Hint: The net effect of the time change on the picture of W is to remove the negative excursions. Modify the excursion measure N accordingly.

6.33 *Processes with two states.* Fix $b > 0$, Let

$$H_t = L_t^\circ + L_t^b,$$

where $L^0 = L$ is the local time at 0, and L^b at b. Show that $X = W_\tau$ is a process with only two states, 0 and b.

a) Show that its jump times form a Poisson process with rate 1/2b, and that the successive jump sizes are $+b,\ -b,\ +b,\ -b, \ldots$. Hint: use proposition 6.19.

b) Compute

$$p_t(x,y) = \mathbb{P}\left\{X_{s+t} = y | X_s = x\right\}$$

for $x,\ y \in \{0,\ b\}$.

6.34 *Processes with three states.* Let $a < 0 < b$ be fixed. Put

$$H_t = L_t^\circ + L_t^a + L_t^b.$$

Show that X is a Markor process whose state space is $D = \{0, a, b\}$. Of course, $X_0 = 0$. Show that the successive states visited by X is a Markov chain $(Y_n)_{n\in\mathbb{N}}$ with $Y_0 = 0$ and transition probability matrix (with states ordered as 0, a, b)

$$P = \begin{bmatrix} 0 & q & p \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

where $p = \mathbb{P}\{Y_{n+1} = b | Y_n = 0\} = -a/(-a+b)$, and $q = 1 - p$. Describe X completely by specifying the distributions of

$$\mathbb{P}\left\{R_n \in dt | Y_n = x\right\}, \qquad x \in D,$$

for the time R_n between the nth and $(n+1)^{th}$ jumps (which is the sojourn time in Y_n).

6.35 *Process on the integers.* Recall that $\mathbb{Z}$ is the set of all integers, positive and negative. Define

$$H_t = \sum_{x\in\mathbb{Z}} L_t^x.$$

For fixed t and ω, show that $H_t(\omega)$ is in fact a finite sum of finite quantities; so, $H_t < \infty$ almost surely for all t in $\mathbb{R}_+$. Show that $X = W_\tau$ is a compound Poisson process, whose jump times form a Poisson process with rate 1 and whose every jump has size ± 1 with equal probabilities.

6.36 *Brownian motion with sticky* 0. Let

$$H_t = t + L_t.$$

The process X is Markov with state space $\mathbb{R}$. It goes through the same states as W does, and in the same order. Show that

$$\int_0^u ds\ 1_{\{0\}} \circ X_s = L_{\tau_u},$$

which is strictly positive for $u > 0$ and increases to $+\infty$ as $u \to \infty$. Describe a method for recovering the path $W(\omega)$ given the path $X(\omega)$.

6.37 *Distribution of AU.* Let A and U be independent, A with the arcsine distribution, and U the uniform on (0,1). Let X, Y, Z be independent gamma variables with shape indices $1/2$, $1/2$, 1 respectively, and with the same scale parameter. Show that

$$A = \frac{X}{X+Y}, \qquad U = \frac{X+Y}{X+Y+Z}$$

satisfy the assumptions on A and U. Conclude that AU has the beta distribution with the index pair $(1/2, 3/2)$.

6.38 *Joint distribution* of G *and* A_G. Write $G = G_1$. In the notation of the proof of Theorem 6.22, similar to 6.27, we can write

$$f(G, A_G) = \int_{\mathbb{R}_+ \times E} N\,(da, dx)\;\; h\left(S^+_{a-}, S^-_{a-}, x\right)$$

where

$$h(u, v, x) = f(u+v, u) 1_{[0,1]}(u+v) 1_{(1,\infty)}(u+v+\zeta(x)).$$

Show that

$$\mathbb{E}\; f(G, A_G) = \int_0^1 ds \frac{1}{\pi\sqrt{s(1-s)}} \int_0^1 dr \quad f(s, sr).$$

Thus, G has the arcsine distribution (as we already know from 2.18); and, given G, the variable A_G has the uniform distribution on $[0,\, G]$.

6.39 *Occupation times.* For t in $\mathbb{R}_+$ and r in $\mathbb{R}$, define

$$A_t(r) = \int_0^t ds\; 1_{(-\infty, r]} \circ W_s.$$

By Theorem 6.22, then, $t - A_t(0) \approx tA$, where A has the arcsine distribution as before. Show that

$$A_t(r) \approx t\; A_1\left(\frac{r}{\sqrt{t}}\right), \qquad A_1(-r) \approx 1 - A_1(r).$$

In view of these, it is enough to concentrate on $A(r) = A_1(r)$ for $r > 0$.

6.40 *Distribution* of $A(r)$. Fix $r > 0$. Show that

$$A(r, \omega) = \begin{cases} 1 & \text{if} \quad T_r(\omega) \geq 1, \\ T_r(\omega) + A_{1-T_r(\omega)}\left(\theta_{T_r(\omega)}\omega\right) & \text{if} \quad T_r(\omega) < 1. \end{cases}$$

Show that, with A independent of T_r,

$$A(r) \approx 1_{\{T_r \geq 1\}} + [T_r + (1+T_r)A]\, 1_{\{T_r < 1\}}.$$

Conclude that, for $u \le 1$,

$$\mathbb{P}\{A(r) < u\} = \int_0^u ds \frac{re^{-r^2/2s}}{\sqrt{2\pi\ s^3}} \mathbb{P}\{s + (1-s)A < u\}.$$

6.41 *Continuation.* This is mere calculus. For $u < 1$, show that

$$\mathbb{P}\{A(r) \in du\} = du \int_0^u ds\ \frac{re^{-r^2/2s}}{\sqrt{2\pi s^3}} \frac{1}{\pi\sqrt{(1-u)(u-s)}}$$

$$= du \frac{1}{\pi\sqrt{u(1-u)}} \int_0^u ds \frac{re^{-r^2/2s}}{\sqrt{2\pi s^3 (1-s/u)}} = du\ \frac{e^{-r^2/2u}}{\pi\sqrt{u(1-u)}}.$$

Hint: In the last integral, replace s with $u/(1+v)$; the integral becomes

$$\int_0^\infty dv\ \frac{re^{-r^2(1+v)/2u}}{\sqrt{2\pi uv}} = e^{-r^2/2u} \int_0^\infty dv \frac{e^{-cv} c^a v^{a-1}}{r(a)} = e^{-r^2/2u},$$

where we recognize the gamma density with $a = 1/2$ and $c = r^2/2u$.

6.42 *Continuation.* To sum up, with Z standard Gaussian, show that

$$\mathbb{P}\{A(r) \in du\} = du \frac{e^{-r^2/2u}}{\pi\sqrt{u(1-u)}} 1_{(0,1)}(u) + \delta_1(du)\mathbb{P}\{|Z| \le r\}.$$

6.43 *Gamma tails and Laplace transforms.* It follows from the preceding computation that

$$\mathbb{E}\ e^{-r^2/2A} = \int_0^1 du \frac{e^{-r^2/2u}}{\pi\sqrt{u(1-u)}} = \mathbb{P}\{|Z| > r\}.$$

Taking $r = \sqrt{2p}$ and recalling that $\frac{1}{2}\ Z^2$ has the standard gamma distribution with shape index $\frac{1}{2}$, we obtain

$$\mathbb{E}\ e^{-p/A} = \mathbb{P}\left\{\frac{1}{2} Z^2 > p\right\} = \int_p^\infty dy \frac{e^{-y}\ y^{-1/2}}{\Gamma(1/2)}.$$

In other words, the tail of the gamma distribution with shape index $1/2$ is the Laplace transform of $1/A$, where A has the arcsine distribution.

6.44 *Brownian Quantiles.* The mapping $r \mapsto A(r)$ is the cumulative distribution function of a random probability measure on $\mathbb{R}$. We define the corresponding quantile function by

$$Q_u = \inf\{r \in \mathbb{R} : A(r) > u\}, \qquad 0 < u < 1.$$

Obviously, $\{Q_u > r\} = \{A(r) < u\}$, and the probabilities of these events can be obtained by using the results of Exercises 6.41–6.42. In particular, for $r > 0$, show that

$$\mathbb{P}\{Q_u \in dr\} = \int_0^u dv \frac{r\ e^{-r^2/2v}}{\pi\sqrt{v^3(1-v)}}.$$

6.45 *Continuation.* It is possible to give a simpler formula for the preceding expression: Since $u < 1$, and $r > 0$, in view of 6.40,

$$\begin{aligned}\mathbb{P}\{Q_u > r\} &= \mathbb{P}\{A + T_r \cdot (1-A) < u\} \\ &= \mathbb{P}\left\{A < u,\ T_r < \frac{u-A}{1-A}\right\} \\ &= \int_0^u dv \frac{1}{\pi\sqrt{v(1-v)}} \int_r^\infty dz\ \frac{2\ e^{-z^2(1-v)/2(u-v)}}{\sqrt{2\pi(u-v)/(1-v)}}\end{aligned}$$

since $T_r \approx r^2/Z^2$, with Z standard Gaussian. Some elementary operations give

$$\begin{aligned}\mathbb{P}\{Q_u \in dr\} &= \frac{2}{\sqrt{2\pi}} \int_0^u dv \frac{1}{\pi\sqrt{v(u-v)}} \exp_- \frac{r^2}{2}\frac{1-v}{u-v} \\ &= \frac{2}{\sqrt{2\pi}} \int_0^1 dx \frac{1}{\pi\sqrt{x(1-x)}} \exp_- \frac{r^2}{2}\left(1 + \left(\frac{1-u}{u}\right)\cdot\frac{1}{v}\right) \\ &= \frac{2}{\sqrt{2\pi}} e^{-r^2/2}\ \mathbb{E}\ \exp_- \frac{r^2(1-u)}{2u}\cdot\frac{1}{A}.\end{aligned}$$

The last expectation can be evaluated using 6.43 to obtain

$$\mathbb{P}\{Q_u \in dr\} = \frac{2}{\sqrt{2\pi}} e^{-r^2/2}\ \mathbb{P}\left\{|Z| > r\sqrt{\frac{1-u}{u}}\right\}.$$

7 Path Properties

This section is on the oscillatory behavior of Brownian paths. We shall see that, for almost every ω, the following are true for the Wiener path $t \mapsto W_t(\omega)$: The path is continuous, but nowhere differentiable. Over every interval, it has infinite total variation, but finite quadratic variation; thus, the path is highly oscillatory, but the oscillations are of small amplitude. In addition to clarifying these points, we shall discuss Hölder continuity of the paths, describe the exact modulus of continuity, and give the law of the iterated logarithm. These help to visualize the paths locally in terms of deterministic functions.

Throughout this section, W is a Wiener process over some probability space $(\Omega, \mathcal{H}, \mathbb{P})$. We assume that the path $W(\omega) : t \mapsto W_t(\omega)$ is continuous for every ω. By a subdivision of an interval $[a, b]$ we mean a finite collection of disjoint intervals of the form $(s, t]$ whose union is $(a, b]$; it is a partition of $(a, b]$ whose elements are intervals. If $\mathcal{A}$ is a subdivision, we write $\|\mathcal{A}\|$ for its *mesh*, defined as $\|\mathcal{A}\| = sup\{t - s : (s, t] \in \mathcal{A}\}$.

Quadratic variation

Let $f: \mathbb{R}_+ \mapsto \mathbb{R}$ be right-continuous. Fix an interval $[a, b]$ in $\mathbb{R}_+$. For $p > 0$ and $\mathcal{A}$ a subdivision of $[a, b]$, consider

7.1
$$\sum_{(s,t]\in\mathcal{A}} |f(t) - f(s)|^p \, .$$

The supremum of this over all such subdivisions $\mathcal{A}$ is called the *true p-variation* of f over $[a, b]$. For $p = 1$, the supremum is called the *total variation* of f on $[a, b]$, and for $p = 2$ the *true quadratic variation.*

These deterministic concepts prove to be too strict when applied to a typical Wiener path: for almost every ω, if $f = W(\omega)$, the total variation over $[a, b]$ is $+\infty$, and so is the true quadratic variation. However, at least for the quadratic variation, a probabilistic version proves interesting:

7.2 Theorem. *Let the interval $[a, b]$ be fixed. Let $(\mathcal{A}_n)$ be a sequence of subdivisions of it with $\|\mathcal{A}_n\| \to 0$. Then, the sequence of random variables*

7.3
$$V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2$$

converges in L^2 and in probability to the length $b - a$.

Proof. Recall that $|W_t - W_s|^2$ has the same distribution as $(t - s)\ Z^2$, where Z is standard Gaussian, and that $\mathbb{E}\ Z^2 = 1,\ \text{var}\ Z^2 = 2$. Since the intervals $(s, t]$ in $\mathcal{A}_n$ are disjoint, the corresponding increments $W_t - W_s$ are independent. Thus,

$$\mathbb{E}\, V_n = \sum_{(s,t]\in\mathcal{A}_n} (t - s) = b - a,$$

$$\text{Var}\ V_n = \sum_{(s,t]\in\mathcal{A}_n} (t - s^2) \cdot 2 \leq 2 \cdot (b - a) \cdot \|\mathcal{A}_n\|.$$

Hence, $\mathbb{E}|V_n - (b - a)|^2 = \text{Var}\ V_n \to 0$ as $n \to \infty$. This shows the convergence in L^2 and implies the convergence in probability. □

The limit in the preceding theorem is called the *quadratic variation* of W over $[a, b]$. Heuristically, it is a sum of squares of the increments over infinitesimal subintervals. The following clarifies this picture by taking the limit for each ω separately.

7.4 Proposition. *For each n in $\mathbb{N}$, let $\mathcal{A}_n$ be the subdivision of $[a, b]$ that consists of 2^n intervals of the same length. Then, (V_n) defined by 7.3 converges to $b - a$ almost surely.*

Proof. Since each $(s,t]$ in $\mathcal{A}_n$ has length $(b-a)\cdot 2^{-n}$, we have $\mathbb{E}\, V_n = b-a$ as before, but Var $V_n = 2^n \cdot 2 \cdot (b-a)^2 \cdot 2^{-2n}$. Thus, Chebyshev's inequality yields that, for $\varepsilon > 0$,

$$\mathbb{P}\left\{|V_n - (b-a)| > \varepsilon\right\} \le \frac{1}{\varepsilon^2} \cdot 2 \cdot (b-a)^2 \cdot 2^{-n}.$$

Since the right side is summable in n, Borel-Cantelli lemma I.2.6 applies, and $V_n \to b\text{-}a$ almost surely. □

7.5 REMARK. a) The preceding proposition can be strengthened. The conclusion remains true when $(\mathcal{A}_n)$ is an arbitrary nested sequence with $\|\mathcal{A}_n\| \to 0$, the term *nested* meaning that each interval of the subdivision $\mathcal{A}_{n+1}$ is a subset of some interval in $\mathcal{A}_n$.

b) But, it is essential that $(\mathcal{A}_n)$ be chosen deterministically. Otherwise, there are counter-examples. For example, for almost every ω there is a nested sequence $(\mathcal{A}_n(\omega))$ with $\|\mathcal{A}_n(\omega)\| \to 0$ such that $V_n(\omega)$ defined by 7.3 goes to $+\infty$.

Total variation

The following proposition shows that each typical path is highly oscillatory over every interval. But the amplitudes must be small enough that their squares sum to the finite number called the quadratic variation.

7.6 PROPOSITION. *For almost every ω, the path $W(\omega)$ has infinite total variation over every interval $[a,\, b]$ with $a < b$.*

Proof. In the setting of Proposition 7.4, let Ω_{ab} be the almost sure set of convergence. Pick ω in Ω_{ab}, write w for $W(\omega)$, and let $v^* \le +\infty$ be the total variation of w over $[a,b]$. We observe that, with sums and supremum over all $(s,t]$ in $\mathcal{A}_n$,

$$\sum |w_t - w_s|^2 \le (\sup |w_t - w_s|) \sum |w_t - w_s| \le (\sup |w_t - w_s|) \cdot v^*,$$

the last inequality being by the definition of v^* as the supremum over all subdivisions. Now, let $n \to \infty$. The left side goes to $b-a \ne 0$ by the way ω is picked. On the right side, the supremum goes to 0 by the uniform continuity of w on $[a,b]$. It follows that v^* cannot be finite.

Let Ω_0 be the intersection of Ω_{ab} over all rationals a and b with $0 \le a < b$. The claim of the proposition holds for $W(\omega)$ for every ω in the almost sure event Ω_0. □

Hölder continuity, nowhere differentiability

Let $\alpha \in \mathbb{R}_+$, $B \subset \mathbb{R}_+$, and $f : \mathbb{R}_+ \mapsto \mathbb{R}$. The function f is said to be *Hölder continuous* of order α on B if there is a constant k such that

7.7 $$|f(t) - f(s)| \le k \cdot |t-s|^{\alpha} \qquad \text{if } s,t \in B.$$

It is said to be locally Hölder continuous of order α if it is such on $[0, b]$ for every $b < \infty$. Note that if f is differentiable at some point then it is Hölder continuous of order 1 on some neighborhood of that point.

The next proposition is another consequence of the finiteness of the quadratic variation, Proposition 7.4. Its proof is similar to that of Proposition 7.6.

7.8 PROPOSITION. *For almost every ω, the Wiener path $W(\omega)$ is Hölder continuous of order α on no interval for α >1/2. In particular, for almost every ω, the path is nowhere differentiable.*

Proof. Pick ω, write w for $W(\omega)$, and suppose that

$$|w_t - w_s| \leq k \cdot |t-s|^\alpha$$

for all s and t in some interval $[a, b]$, $a < b$, for some $\alpha > 1/2$ and some constant k. With $\mathcal{A}_n$ as in Proposition 7.4, with summations and supremum over $(s, t]$ in $\mathcal{A}_n$,

$$\sum |w_t - w_s|^2 \leq k^2 \sum |t-s|^{2\alpha} \leq k^2 \cdot (b-a) \cdot \sup |t-s|^{2\alpha - 1}.$$

As $n \to \infty$, the supremum vanishes since $2\alpha > 1$, which means that the left side vanishes as well. Thus, ω does not belong to the almost sure set Ω_{ab} of convergence in Proposition 7.4. Hence, the claims hold for every ω in the intersection of Ω_{ab} over all rationals $a < b$. □

7.9 REMARK. We shall see shortly that the claim of the preceding proposition remains true for $\alpha = 1/2$ as well; see Theorem 7.13 below.

By contrast, the following is a positive result. Its proof is based on a lemma of independent interest; the lemma is put last in this section in order to preserve the continuity of presentation; see 7.32.

7.10 PROPOSITION. *For almost every ω, the path $W(\omega)$ is locally Hölder continuous of order α for every $\alpha < 1/2$.*

Proof. For Z standard Gaussian, $c_p = \mathbb{E}\, Z^{2p} < \infty$, and

$$\mathbb{E}\, |W_t - W_s|^{2p} = c_p\, |t-s|^p, \qquad p \geq 1.$$

Thus, Lemma 7.32 below applies, and almost every path is Hölder continuous of order $\alpha = (p-1)/2p = 1/2 - 1/2p$ on [0,1]. Scaling property allows us to replace [0,1] with [0,b] for each $b < \infty$, and the proof is complete since p can be taken as large as desired. □

Modulus of continuity

Let f and g be functions from [0,1] into $\mathbb{R}$. The function g is said to be a *modulus of continuity* for f if

7.11 $$s, t \in [0,1]\,,\; [t-s] \leq \delta \Rightarrow |f(t) - f(s)| \leq g(\delta)$$

for every $\delta > 0$ small enough. Of course, then, so is cg for every constant $c \geq 1$. The following theorem, due to Lévy, shows that

7.12 $$g(t) = \sqrt{2t\log(1/t)}, \qquad t \in [0,1],$$

is the exact modulus of continuity for the paths of $(W_t)_{t\in[0,1]}$ in the following sense: cg is a modulus of continuity for almost every path if $c > 1$, and is a modulus of continuity for almost no path if $c < 1$. The proof will be delayed somewhat; see 7.26.

7.13 Theorem. *Let g be as defined by 7.12. Then, for almost every ω,*

$$\limsup_{\delta\to 0} \frac{1}{g(\delta)} \sup_{\substack{0 \leq s < t \leq 1 \\ t - s \leq \delta}} |W_t(\omega) - W_s(\omega)| = 1.$$

As a corollary, since $\sqrt{\delta}/g(\delta)$ goes to 0 as δ goes to 0, we obtain the proof of Remark 7.9. Details are left as an exercise.

Law of the iterated logarithm

This is about the oscillatory behavior of Wiener paths near the time 0 and for very large times. The name comes from its use of

7.14 $$h(t) = \sqrt{2t\;\log\log(1/t)}, \qquad t \in [0,1],$$

as the control function.

7.15 Theorem. *With h as in 7.14, the following hold for almost every ω:*

$$\limsup_{t\to 0} \frac{1}{h(t)} W_t(\omega) = 1, \qquad \liminf_{t\to 0} \frac{1}{h(t)} W_t(\omega) = -1.$$

7.16 Remark. By time inversion, the same results hold when $W_t(\omega)$ is replaced with $t\,W_{1/t}(\omega)$. Then, replacing $1/t$ with t, we obtain that the following hold for almost every ω:

7.17 $$\limsup_{t\to\infty} \frac{W_t(\omega)}{\sqrt{2t\log\log t}} = 1, \quad \liminf_{t\to\infty} \frac{W_t(\omega)}{\sqrt{2t\log\log t}} = -1.$$

We know from Lemma 3.3 that the running maximum increases to $+\infty$, and the running minimum decreases to $-\infty$ in the limit as $t \to \infty$. These are re-confirmed by 7.17 and are made more precise.

Proofs. We list here two approximation lemmas before giving the proofs of the last two theorems.

7.18 LEMMA. $\mathbb{P}\left\{\sup_{t\le 1}\left(W_t - \tfrac{1}{2}\, pt\right) > q\right\} \le e^{-pq}$ *for positive* p *and* q.

Proof. Let $X_t = \exp\left(pW_t - \tfrac{1}{2}\, p^2 t\right)$. The probability in question is

$$\mathbb{P}\left\{\sup_{t\le 1}\ X_t > e^{pq}\right\} \le e^{-pq}\ \mathbb{E}\ X_1 = e^{-pq},$$

where the inequality follows from the maximal inequality V.5.33 applied to the exponential martingale X. □

7.19 LEMMA. *Let* Z *be a standard Gaussian variable. Then, for* $b > 0$,

$$\frac{1}{4}\cdot\frac{b}{1+b^2}\ e^{-b^2/2} < \mathbb{P}\{Z > b\} < \frac{1}{2b}e^{-b^2/2}.$$

Proof. Observe that

$$\int_b^\infty dx\ e^{-x^2/2} < \int_b^\infty dx\ \frac{x}{b}\ e^{-x^2/2} = \frac{1}{b}\ e^{-b^2/2},$$

and

$$\int_b^\infty dx\ e^{-x^2/2} > \int_b^\infty dx\ \frac{b^2}{x^2}\ e^{-x^2/2} = b\ e^{-b^2/2} - b^2\int_b^\infty dx\ e^{-x^2/2},$$

the last equality being through integration by parts. The rest is arithmetic. □

7.20 *Proof of Theorem* 7.15. a) We show first that, for almost every ω,

7.21 $$\alpha(\omega) = \limsup_{t\to 0}\frac{1}{h(t)}W_t(\omega)$$

is at most 1.

Let $0 < a < 1 < b$. Put $p_n = b\ h(a^n)/a^n$ and $q_n = \tfrac{1}{2}\ h(a^n)$. By Lemma 7.18,

$$\mathbb{P}\left\{\sup_{t\le 1}\left(W_t - \tfrac{1}{2}\ p_n t\right) > q_n\right\} \le e^{-p_n q_n} = \left(n\log \tfrac{1}{a}\right)^{-b},$$

and the right side is summable in n. Thus, by the Borel-Cantelli lemma, there is an almost sure event Ω_{ab} such that for every ω in it there is n_ω such that

7.22 $$W_t(\omega) \le q_n + \tfrac{1}{2}\ p_n t \qquad \text{for every } t \le 1 \text{ and } n \ge n_\omega.$$

The function h is increasing on $[0, e^{-c}]$, where $c = e^{1/c}$. Choose $n \ge n_\omega$ large enough that $a^{n-1} \le e^{-c}$, and let $t \in (a^n,\ a^{n-1}]$. By 7.22,

$$W_t(\omega) \le q_n + \tfrac{1}{2}p_n\ a^{n-1} = \frac{1}{2}\cdot\left(1+\frac{b}{a}\right)h(a^n) \le \frac{1}{2}\left(1+\frac{b}{a}\right)h(t).$$

Hence, for ω in Ω_{ab},

7.23
$$\alpha(\omega) \leq \frac{1}{2}\left(1+\frac{b}{a}\right).$$

write Ω_n for Ω_{ab} with $a = 1 - (1/n)$ and $b = 1 + (1/n)$. Now 7.23 implies that $\alpha(\omega) \leq 1$ for every ω in the almost sure event $\cap_n \Omega_n$.

b) Next, we prove that $\alpha(\omega) \geq 1$ for almost every ω. Let $\varepsilon \in (0, 1)$ and put $t_n = \varepsilon^{2n}$. Observe that $h(t_{n+1}) \leq 2\ \varepsilon\ h(t_n)$ for all n large enough. And, by part(a) applied to the Wiener process $(-W_t)$, there is an almost sure set Ω_o such that 7.21 holds, with $-W$ replacing W, for almost every ω. Thus,

7.24
$$\omega \in \Omega_0 \Rightarrow -W_{t_{n+1}}(\omega) \leq 2\ h(t_{n+1}) \leq 4\ \varepsilon\ h(t_n) \quad \text{for all } n \text{ large enough.}$$

On the other hand, by Lemma 7.19 applied with $b = (1-\varepsilon)h(t_n)/\sqrt{t_n - t_{n+1}}$,

$$p_n = \mathbb{P}\left\{W_{t_n} - W_{t_{n+1}} > (1-\varepsilon)h(t_n)\right\} > \frac{1}{4}\,\frac{b}{1+b^2}\ e^{-b^2/2},$$

and $e^{-b^2/2} = (2\ \ n\ \ \log\ \ 1/\varepsilon)^{-c}$, where $c = (1-\varepsilon)/(1+\varepsilon)$ is less than 1. It follows that $\Sigma_{p_n} = +\infty$. Since the increments $W_{t_n} - W_{t_{n+1}}$ are independent, the divergence part of the Borel-Cantelli lemma applies. There is an almost sure event Ω_ε such that

7.25 $\omega \in \Omega_\varepsilon \Rightarrow W_{t_n}(\omega) - W_{t_{n\pm 1}}(\omega) > (1-\varepsilon)\ h(t_n) \quad$ for infinitely many n.

Combining 7.24 and 7.25, we see that

$$\begin{aligned} \omega \in \Omega_0 \cap \Omega_\varepsilon &\Rightarrow W_{t_n}(\omega) > (1-5\varepsilon)\ h(t_n) \\ &\Rightarrow \limsup_{t\to 0} \frac{1}{h(t)} W_t(\omega) \geq \limsup_{n\to\infty} \frac{1}{h(t_n)} W_{t_n}(\omega) \geq 1 - 5\varepsilon. \end{aligned}$$

For $k \geq 1$, put $\Omega_k = \Omega_o \wedge \Omega_\varepsilon$ with $\varepsilon = 1/k$. Then, for ω in $\bigcap \Omega_k$,

$$\limsup_{t\to 0} \frac{1}{h(t)} W_t(\omega) \geq 1.$$

This completes the proof of the statement about the limit superior. The one about the limit inferior is obtained by recalling that lim inf $x_n = -$ lim sup$(-x_n)$ and that $-W$ is again Wiener.

7.26 *Proof of Theorem* 7.13. a) First we show that

7.27
$$\alpha(\omega) = \limsup_{\delta\to 0} \frac{1}{g(\delta)} \sup_{\substack{0\,\leq\, s\,<\,t\,\leq\, 1 \\ t\,-\,s\,\leq\,\delta}} |W_t(\omega) - W_s(\omega)|$$

is equal to 1 or more for every ω in an almost sure event Ω_o.

Take a in (0,1), put $u = 2^{-n}$, and recall g from 7.12. Note that $g(u)/\sqrt{u} = b\sqrt{n}$, where $b = \sqrt{2\log 2}$, and $e^{-nb^2/2} = 2^{-n}$. For Z standard Gaussian, it follows from Lemma 7.19 that

$$p = \mathbb{P}\left\{\sqrt{u}\ |Z| > a\ g(u)\right\} > \frac{1}{2}\frac{ab\sqrt{n}}{1+a^2b^2n}\ e^{-a^2b^2n/2} > c\ 2^{-na^2}/\sqrt{n},$$

for some constant c depending on a only. Thus, since the increments $W_{ku} - W_{ku-u}$ are independent and identically distributed as $\sqrt{u}$ Z,

$$\mathbb{P}\left\{\max_{1\le k\le 2^n} |W_{ku} - W_{ku-u}| \le a\ g(u)\right\}$$
$$= (1-p)^{2^n} \le e^{-p2^n} \le \exp\left(-c2^{n-na^2}/\sqrt{n}\right)$$

since $1 - p \le e^{-p}$. The right-most member is summable in n. Hence, by the Borel-Cantelli lemma, there is an almost sure event Ω_a such that

$$\omega \in \Omega_a, u = 2^{-n} \Rightarrow \max_{1\le k\le 2^n} |W_{ku}(\omega) - W_{ku-u}(\omega)| > a\ g(u)$$

for all n large enough, which means that $\alpha(\omega) > a$; see 7.27 for $\alpha(\omega)$. Let Ω_0 be the intersection of Ω_a over a in $\{1/2,\ 2/3,\ 3/4,\ \ldots\}$; then, Ω_0 is almost sure, and $\alpha(\omega) \ge 1$ for ω in Ω_0.

b) We show next that $\alpha(\omega) \le 1$ for almost every ω. Choose $b > 1$. Put $a = 2/(1+b)$. Note that $a \in (0,1)$ and $ab > 1$. For u in $(0,\ 2^{-na})$, since $g(u)/\sqrt{u} \ge \sqrt{2\,na\,\log 2}$, it follows from Lemma 7.19 that

$$\mathbb{P}\left\{\sqrt{u}\,|Z| > b\,g(u)\right\} \le \mathbb{P}\left\{|Z| > b\sqrt{2\,na\,\log 2}\right\}$$

7.28 $$\le \frac{e^{-b^2\,na\,\log 2}}{b\sqrt{2\,na\,\log 2}} = c\cdot 2^{-na\,b^2}/\sqrt{n},$$

where c depends only on b.

Let B_n be the set of all pairs of numbers s and t in the set $D_n = \{k/2^n : 0 \le k \le 2^n\}$ satisfying $0 < t - s < 2^{-na}$; there are at most 2na such pairs (s,t). Using 7.28 with $u = t - s$, we get

$$\mathbb{P}\left\{\max_{(s,t)\in B_n} \frac{1}{g(t-s)}\,|W_t - W_s| > b\right\} \le 2^{na}\cdot c\cdot 2^{-na\,b^2}/\sqrt{n};$$

and the right side is summable in n, since $ab^2 - a > b - a > 0$. Thus, by the Borel-Cantelli lemma, there is an almost sure event Ω_b such that for every ω in it there is n_ω such that

7.29 $$n \ge n_\omega, (s,t) \in B_n \Rightarrow |W_t(\omega) - W_s(\omega)| \le b\ g(t-s).$$

Fix ω in Ω_b; write n^* for n_ω, and w for $W(\omega)$. Let $D = \bigcup_0^\infty D_m$, the set of all dyadic numbers in [0,1]. For s and t in D, put $s_m = \inf D_m \cap [s,1]$ and

$t_m = sup\ D_m \cap [0, t]$. Then, (s_m) is decreasing, (t_m) is increasing, and $s_m = s$ and $t_m = t$ for all m large enough. Thus,

7.30 $$w_t - w_s = \sum_{m \geq n} \left(w_{t_{m+1}} - w_{t_m} \right) + w_{t_n} - w_{s_n} + \sum_{m \geq n} \left(w_{s_m} - w_{s_{m+1}} \right).$$

Suppose that $0 < t - s < 2^{-n^* a}$ and choose $n \geq n^*$ such that

7.31 $$2^{-na-a} \leq t - s < 2^{-na} < e^{-1}.$$

Then, $s \leq s_n \leq t_n \leq t$, and the times $t_{m+1},\ t_m,\ s_m,\ s_{m+1}$ belong to B_{m+1} for every $m \geq n$. It follows from 7.29 and 7.30 that

$$|w_t - w_s| \leq \sum_{m \geq n} b\ \ g(t_{m+1} - t_m) + b\ \ g(t-s) + \sum_{m \geq n} b\ \ g\left(s_m - s_{m+1}\right).$$

Moreover, g is increasing on $[0, e^{-1}]$, and $t_{m+1} - t_m \leq 2^{-m-1} \leq e^{-1}$ and $s_m - s_{m+1} \leq 2^{-m-1} \leq e^{-1}$ for $m \geq n$ by the way n is chosen. So,

$$|w_t - w_s| \leq b\ \ g(t-s) + 2b \sum_{m \geq n} g(2^{-m-1}).$$

Also,

$$\sum_{m>n} g(2^{-m}) = g(2^{-n}) \sum_{m>n} \sqrt{2^{-m+n} m/n} \leq g(2^{-n}) \sum_{p \geq 1} \sqrt{2p\, 2^{-p}}$$

and

$$g\left(2^{-n}\right) \leq g(t-s) g(2^{-n})/g(2^{-na-a}) \leq g(t-s) \sqrt{2^{-n(1-a)} \cdot 2^a / a}$$

in view of 7.31. Combining the last three expressions, we see that

$$|w_t - w_s| \leq b\ g(t-s) + 2bc\ g(t-s) \sqrt{2^{-n(1-a)}}$$

with c chosen appropriately. This was for s and t in D satisfying 7.31; by the continuity of w, the same holds for all s and t in [0,1] satisfying 7.31. Consequently, letting $n \to \infty$ and recalling that $1 - a > 0$, we see for $\alpha(\omega)$ of 7.27 that $\alpha(\omega) \leq b$ for the arbitrarily fixed ω in Ω_b. Thus, $\alpha(\omega) \leq 1$ for every ω in $\bigcap \Omega_b$, where the intersection is over b in $\{1 + 1/n : n \geq 1\}$. This completes the proof. □

Kolmogorov's moment condition

The next lemma was used to prove Hölder continuity in Proposition 7.10. It is the main part of Kolmogorov's theorem on the existence of continuous modifications. These are stated in a form that will be of use in the next section. Recall that D is the set of all dyadic numbers in [0,1]. Here, $X = (X_t)_{t \in [0,1]}$ is a process with state space $\mathbb{R}$.

7.32 LEMMA. *Suppose that there exist constants c, p, q in (0,1) such that*

7.33 $$\mathbb{E}\,|X_t - X_s|^p \le c \cdot |t-s|^{1+q}, \qquad s,t \in [0,1].$$

Then, for every α in $[0,q/p)$ there is a random variable K such that $\mathbb{E}\,K^p$ is finite and

7.34 $$|X_t - X_s| \le K \cdot |t-s|^{\alpha}, \qquad s,t \in D.$$

If X is also continuous, then 7.34 holds for all s, t in [0,1].

Proof. Fix α in $[0, q/p)$. Let

7.35 $$K = \sup_{s,t\in D,\ s\neq t} \frac{|X_t - X_s|}{|t-s|^{\alpha}}.$$

Since $D \times D$ is countable, this defines a random variable. Now, 7.34 is obvious, and it extends to s, t in [0,1] when X is continuous, since D is dense in [0,1]. Thus, the proof reduces to showing that

7.36 $$\mathbb{E} \quad K^p < \infty.$$

a) Let $M_n = \sup |X_t - X_s|$, where the supremum is over all pairs of numbers s and t in D_n with $t - s = 2^{-n}$. Since there are 2^n such pairs, the assumption 7.33 implies that

7.37 $$\mathbb{E}\,M_n^p \le 2^n \cdot c \cdot (2^{-n})^{1+q} = c \cdot 2^{-nq}.$$

b) For s and t in D, let $s_n = \inf\ D_n \cap [s,1]$ and $t_n = \sup\ D_n \cap [0,t]$. Then, (s_n) is decreasing, (t_n) is increasing, and $s_n = s$ and $t_n = t$ for all n large enough. Thus,

$$X_t - X_s = \sum_{n\ge m} \left(X_{t_{n+1}} - X_{t_n}\right) + X_{t_m} - X_{s_m} + \sum_{n\ge m} \left(X_{s_n} - X_{s_{n+1}}\right).$$

If $0 < t - s \le 2^{-m}$, then $t_m - s_m$ is either 0 or equal to 2^{-m}; hence,

7.38 $$|X_t - X_s| \le \sum_{n\ge m} M_{n+1} + M_m + \sum_{n\ge m} M_{n+1} \le 2 \sum_{n\ge m} M_n.$$

c) Consider 7.35. Take the supremum there first over s and t with $2^{-m-1} < |t-s| \le 2^{-m}$ and then over m. In view of 7.38, we get

$$K \le \sup_m \left(2^{m+1}\right)^{\alpha} \cdot 2 \sum_{n\ge m} M_n \le 2^{1+\alpha} \sum_{n\ge 0} 2^{n\alpha} M_n.$$

If $p \ge 1$, letting $\|\cdot\|$ denote the L^p-norm, we see from 7.37 that

$$\|K\| \le 2^{1+\alpha} \sum_n 2^{n\alpha}\ c^{1/p}\ 2^{-nq/p} < \infty$$

since $\alpha < q/p$. If $p < 1$, then $(x+y)^p \le x^p + y^p$ for positive x and y, and

$$\mathbb{E}K^p \le (2^{1+\alpha})^p \sum_n 2^{n\alpha p} c \cdot 2^{-nq} < \infty$$

again. Thus, 7.36 holds in either case, as needed to complete the proof. □

The following is Kolmogorov's theorem on modifications. Recall that $\tilde{X}$ is a modification of X if for every t there is an almost sure event Ω_t on which $\tilde{X}_t = X_t$.

7.39 THEOREM. *Suppose that 7.33 holds for some constants, c, p, q in $(0, \infty)$. Then, for every α in $[0, q/p]$ there is a modification $\tilde{X}$ of X such that the path $\tilde{X}(\omega)$ is Hölder continuous of order α on [0, 1] for every ω.*

Proof. Fix α as described. Let K be as in Lemma 7.32. Since $\mathbb{E}\, K^p < \infty$, the event $\Omega_0 = \{K < \infty\}$ is almost sure. For ω outside Ω_0, put $\tilde{X}(\omega) = 0$ identically. For ω in Ω_0, Lemma 7.32 ensures that $X(\omega)$ is Hölder continuous of order α on D. Thus, putting

7.40 $$\tilde{X}_t(\omega) = \lim_{\substack{s \to t \\ s \in D}} X_s(\omega), \qquad t \in [0,1], \quad \omega \in \Omega_0,$$

we obtain a path $\tilde{X}(\omega)$ that is Hölder continuous of order α on [0,1]. The same property holds trivially for $\tilde{X}(\omega)$ with $\omega \notin \Omega_0$. Finally, for each t in [0,1], we have $X_t = \tilde{X}_t$ almost surely in view of 7.40 and 7.33. □

Exercises

7.41 *p-variation.* Let $\mathcal{A}_n$ be the subdivision of [0,1] that consist of $(0, \delta]$, $(\delta, 2\delta]$, ..., $(1-\delta, 1]$ with $\delta = 1/n$. Show that, for $p > 0$,

$$\frac{1}{n}\sqrt{n^p} \sum_{(s,t] \in \mathcal{A}_n} |W_t - W_s|^p$$

converges, as $n \to \infty$, to $\mathbb{E}|Z|^p$ in probability, where Z is standard Gaussian. Hint: Use time inversion and the weak law of large numbers.

7.42 *Monotonicity.* For almost every ω, the Wiener path is monotone in no interval. Show. Hint: Compare Proposition 7.6 with Exercise I.5.24.

7.43 *Local maxima.* Let $f : [0,1] \mapsto \mathbb{R}$ be continuous. It is said to have a *local maximum* at t if there is $\varepsilon > 0$ such that $f(s) \le f(t)$ for every s in $(t - \varepsilon,\ t + \varepsilon)$. Suppose that f is monotone in no interval. Show the following:

a) f has a local maximum.

b) If f has local maxima at s and at t, then it has a local maximum at some point u in (s, t).

c) The set of all local maxima is dense in [0,1].

7.44 *Exponential inequality.* This is similar to Lemma 7.18. Let $M_t = \max_{s\le t} W_s$. Show that, for $a > 0$,

$$\mathbb{P}\{M_t > at\} \le e^{-a^2 t/2}.$$

Hint: Recall that $M_t \approx \sqrt{t}|Z|$. So, $M_t^2 \approx t\, Z^2 \le t(Z^2 + Y^2) \approx 2tX$, where Y and Z are independent standard Gaussians, and X is standard exponential.

8 Existence

This is to end the chapter by completing the circle, by showing that Brownian motions do exist. The question of existence is mathematical: Does there exist a probability space, and a process defined over it, such that the process is continuous and appropriately Gaussian.

We give two very different constructions. The first is via Kolmogorov's extension theorem and existence of continuous modifications; here, the complexities of the Wiener process are built into the probability measure in an abstract fashion. By contrast, the second, due to Lévy, uses a very simple probability space, and the intricacies of the process are built explicitly into the paths.

First construction

The basic ingredients are Theorem IV.4.18, the Kolmogorov extension theorem, and Theorem 7.39 on the existence of continuous modifications.

8.1 Theorem. *There exist a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and a stochastic process $W = (W_t)_{t\in\mathbb{R}_+}$ such that W is a Wiener process over $(\Omega, \mathcal{H}, \mathbb{P})$.*

Proof. We follow the setup of Theorem IV.4.18. Let Ω be the set of all mappings from $\mathbb{R}_+$ into $\mathbb{R}$. For t in $\mathbb{R}_+$ and ω in Ω, put $X_t(\omega) = \omega(t)$. Let $\mathcal{H}$ be the σ-algebra on Ω generated by $\{X_t : t \in \mathbb{R}_+\}$. For each finite subset J of $\mathbb{R}_+$, if J has $n \ge 1$ elements, let π_J be the n-dimensional Gaussian distribution on $\mathbb{R}^J$ with mean 0 and covariances $s \wedge t$ for s and t in J. These finite-dimensional distributions π_J form a consistent family. Thus, by Theorem IV.4.18, there exists a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{H})$ such that the distribution of $(X_t)_{t\in J}$ under $\mathbb{P}$ is given by π_J for every finite subset J of $\mathbb{R}_+$. It follows that the process $X = (X_t)_{t\in\mathbb{R}_+}$ has stationary and independent increments, has $X_0 = 0$ almost surely, and every increment $X_t - X_s$ has the Gaussian distribution with mean 0 and variance $t - s$.

Consider $(X_t)_{t\in[0,1]}$. Note that the condition 7.33 holds, for instance, with $p = 4$, $q = 1$, $c = 3$. Thus, Theorem 7.39 applies: there is a modification $(\tilde{X}_t)_{t\in[0,1]}$ that is continuous. Applying 7.39 repeatedly to $(X_t)_{t\in[n,n+1]}$, $n \in \mathbb{N}$, we obtain a process $\tilde{X} = (\tilde{X}_t)_{t\in\mathbb{R}_+}$ that is continuous and has the same finite-dimensional distributions as X. Thus, $\tilde{X}$ is the Wiener process W sought. □

Lévy's construction

This starts with a probability space on which there is defined a countable independency of standard Gaussian variables Z_q, one for each q in the set D of all dyadic numbers in [0,1]. This is easy to do; see the next theorem, and also the exercises below which show that the probability space can be taken to be ((0,1), $\mathcal{B}_{(0,1)}$, Leb).

The object is to construct $X = \{X(t) : t \in [0,1]\}$ such that X is a Wiener process on [0,1]. It will be obtained as the limit of a sequence of piecewise linear continuous processes X_n. The initial process is defined as

8.2 $$X_0(t) = t\ Z_1, \qquad t \in [0,1].$$

By the nth step, the variables $X(t)$ will have been defined for t in $D_n = \{k/2^n :\ k = 0,1,\ldots,2^n\}$, and the process $X_n = \{X_n(t) = t \in [0,1]\}$ is the piecewise linear continuous process with $X_n(t) = X(t)$ for t in D_n. At the next step, $X(t)$ is specified for t in $D_{n+1}\backslash D_n$ and X_{n+1} is defined to be the piecewise linear continuous process with $X_{n+1}(t) = X(t)$ for t in D_{n+1}. See Figure 16.

To implement this plan, we need to specify $X(q)$ for q in $D_{n+1}\backslash D_n$ consistent with the values $X(t)$ for t in D_n. This problem is solved by Example 1.9 since X is to be Wiener: Given $X(p)$ and $X(r)$ for adjacent points p and r in D_n, the conditional distribution of $X(q)$ at the midpoint q of $[p,r]$ must be Gaussian with mean $\frac{1}{2}X(p) + \frac{1}{2}X(r)$ and variance 2^{-n-2}; note that the conditional mean is exactly $X_n(q)$; thus, we should put $X(q) = X_{n+1}(q) = X_n(q) + \sqrt{2^{-n-2}}\ Z_q$. Finally, piecewise linearity of X_n and X_{n+1} require that we put

8.3 $$X_{n+1}(t) = X_n(t) + \sum_{q\in D_{n+1}\backslash D_n} h_q(t)\ Z_q, \qquad n \ge 0,\ t \in [0,1],$$

where

8.4 $$h_q(t) = \sqrt{2^{-n-2}}\left(1 - |t-q|\cdot 2^{n+1}\right)^+, \qquad q \in D_{n+1}\backslash D_n.$$

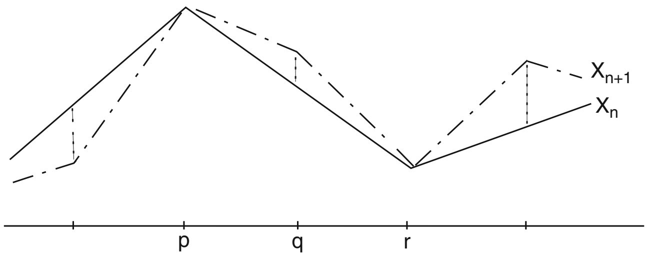


Figure 16: Approximation X_{n+1} coincides with X_n at the points p and r in D_n and differs from X_n at the midpoint q by an amount $Z_q/\sqrt{2^{n+2}}$.

8.5 REMARKS. a) For q in $D_{n+1} \backslash D_n$, the function h_q achieves its maximum $\sqrt{2^{-n-2}}$ at the point q and vanishes outside the interval of length 2^{-n} centered at q. Thus, in particular, for each t, the sum in 8.3 has at most one non-zero term.

b) Note that all the h_q are re-scaled translations of the "mother wavelet" $h(t) = (1 - |t|)^+,\ t \in [-1, 1]$.

The following is the formal construction and the proof that the sequence of process X_n converges to a Wiener process.

8.6 THEOREM. *Let μ be the standard Gaussian distribution on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. *Define*

8.7 $$(\Omega, \mathcal{H}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu)^D,$$

and let Z_q, $q \in D$, be the coordinate variables. Let $X_0, X_1, \ldots$ be defined by 8.2 and 8.3. Then, there exists a process $X = \{X(t) :\ t \in [0, 1]\}$ *such that, for almost every ω in Ω,*

8.8 $$\lim_{n\to\infty} \sup_{t\in[0,1]} |X_n(\omega, t) - X(\omega, t)| = 0;$$

and the process X is a Wiener process with parameter set [0,1].

Proof. a) Existence and construction of the probability space of 8.7 is immediate from Theorem IV.4.7; see IV.5.1 *et seq.* as well. It is clear that $\{Z_q : q \in D\}$ is an independency of standard Gaussian variables.

b) For $f : [0, 1] \mapsto \mathbb{R}$, let $\|f\| = \sup_t |f(t)|$, the supremum norm. We shall show that (X_n) is Cauchy for almost sure convergence in the norm $\|\cdot\|$. This implies the existence of a continuous process X such that $\|X_n - X\| \to 0$ almost surely, and there remains to show that X is Gaussian with mean 0 and covariance $s \wedge t$ for $X(s)$ and $X(t)$. To that end, we observe from 8.2 and 8.3 that $\{X_n(t) :\ t \in D_n\}$ is Gaussian with mean 0 and covariance $s \wedge t$; and the same is true for $\{X_{n+k}(t) :\ t \in D_n\}$ for every k, since $X_{n+k}(t) = X_n(t)$ for $t \in D_n$. Hence, $\{X(t) : t \in D_n\}$ is Gaussian with mean 0 and variance $s \wedge t$, which means that the same is true for $\{X(t) : t \in D\}$. In view of the continuity of X, approximating $X(t)$ by $X(q)$, $q \in D$, we see that X is Gaussian as desired, thus completing the proof.

c) Fix $n \geq 8$. Put $\varepsilon = 2^{-(n+2)/4}$. In view of 8.3 and Remark 8.5a, noting that the maximum of h_q is ε^2, we see that $\|X_{n+1} - X_n\| = \varepsilon^2 M$, where M is the maximum of $|Z_q|$ as q ranges over the set $D_{n+1} \backslash D_n$ of cardinality 2^n. Since the Z_q are independent copies of the standard Gaussian Z_0,

$$\begin{aligned}\mathbb{P}\{\|X_{n+1} - X_n\| > \varepsilon\} &= \mathbb{P}\left\{\varepsilon^2 M > \varepsilon\right\} \\ &\leq 2^n\ \mathbb{P}\left\{|Z_0| > 1/\varepsilon\right\} \leq 2^n \cdot \varepsilon \cdot e^{-1/2\varepsilon^2},\end{aligned}$$

the last inequality being by Lemma 7.19. Since $(1/2\varepsilon^2) = \sqrt{2^n} \geq 2n$ for $n \geq 8$, and since $e^{-2n} \leq 2^{-2n}$, we conclude that, with $\varepsilon_n = 2^{-(n+2)/4}$,

8.9 $$\sum_n \mathbb{P}\{\|X_{n+1} - X_n\| > \varepsilon_n\} < \infty, \qquad \sum_n \varepsilon_n < \infty.$$

By the Borel-Cantelli lemma, then, there exists an almost sure event Ω_0 such that, for every ω in it there is n_ω with

$$\|X_{n+1}(\omega,\cdot) - X_n(\omega,\cdot)\| \le \varepsilon_n \qquad \text{for all} \quad n \ge n_\omega .$$

Thus, for ω in Ω_0, if $i, j \ge n \ge n_\omega$,

$$\|X_i(\omega,\cdot) - X_j \cdot (\omega,\cdot)\,\| \le \sum_{k=n}^{\infty} \varepsilon_k ,$$

and the right side goes to 0 as $n \to \infty$ since (ε_k) is summable. So, for ω in Ω_0, the sequence $(X_n(\omega,\cdot))$ is Cauchy for convergence in the norm and, hence, has a limit $X(\omega,\cdot)$ in the norm. We re-define $X(\omega,\cdot) = 0$ identically for ω not in Ω_0. This X is the process that was shown to be Wiener in part (b) above. □

Exercises

8.10 *Construction on* [0,1] *with its Lebesgue measure.* This is to show that, in Lévy's construction, we can take $(\Omega, \mathcal{H}, \mathbb{P})$ to be $([0,1], \mathcal{B}_{[0,1]}, \text{Leb})$. This is tedious but instructive.

Let $A = \{0,1\}$, $\mathcal{A} = 2^A$, and α the measure that puts weight ½ at the point 0, and ½ at the point 1; then $(A,\ \mathcal{A},\ \alpha)$ is a model for the toss of a fair coin once. Thus,

$$(\Omega, \mathcal{H}, \mathbb{P}) = (A, \mathcal{A}, \alpha)^{\mathbb{N}^*}$$

is a model for an infinite sequence of tosses, independently. We know that $(\Omega,\ \mathcal{H},\ \mathbb{P})$ is basically the same as ([0,1], $\mathcal{B}_{[0,1]}$, Leb).

Let $b:\ \mathbb{N}^* \times \mathbb{N}^* \mapsto \mathbb{N}^*$ be a bijection, and define

$$U_i(\omega) = \sum_{j=1}^{\infty} 2^{-j} \omega_{b(i,j)} \qquad \text{if } \omega = (\omega_1, \omega_2, \cdots) .$$

Show that $U_1,\ U_2,\ \ldots$ are independent and uniformly distributed on [0,1]. Let h be the quantile function (inverse functional) corresponding to the cumulative distribution function for the standard Gaussian. Then, $Y_1 = h \circ U_1,\ Y_2 = h \circ U_2,\ \ldots$ are independent standard Gaussian variables. Finally, let $g:\ D \mapsto \mathbb{N}^*$ be a bijection, and put $Z_q = Y_{g(q)}$ for $q \in D$. Then, $\{Z_q\ :\ q \in D\}$ is an independency of standard Gaussian variables over the probability space $(\Omega, \mathcal{H}, \mathbb{P})$. Lévy's construction yields a Wiener path $W_t(\omega),\ t \in [0,1]$, for each sequence ω of zeros and ones.

88.11 *Lévy's construction, an alternative.* Start with the probability space $(\Omega,\ \mathcal{H},\ \mathbb{P})$ and the standard Gaussians $Z_q,\ q \in D$. Put $W_0 = 0,\ W_1 = Z_1$. Having defined W_p for every p in D_n, put

$$W_q = \frac{1}{2}\,(W_p + W_r) + \sqrt{2^{-n-2}}\; Z_q, \qquad q \in D_{n+1} \backslash D_n ,$$

where $p = \sup D_n \cap [0,q]$ and $r = \inf D_n \cap [q,1]$.

a) Show that $\{W_t : t \in D\}$ is a Gaussian process; specify its mean and covariance function.,

b) Show that the condition 7.33 is satisfied (with $p=4$, $q=1$) for $\{W_t : t \in D\}$ with s, t in D. Then 7.34 holds. Show that this implies that, for almost every ω, the function $t \mapsto W_t(\omega)$ from D into $\mathbb{R}$ is uniformly continuous; let $t \mapsto \bar{W}_t(\omega)$ be its continuous extension onto $[0, 1]$. For the negligible set of ω remaining, put $\bar{W}_t(\omega) = 0$. Show that $\bar{W}$ is a Wiener process on [0,1].

Chapter IX

Markov Processes

A stochastic process is said to have the Markov property if, at every instant, given the past until that instant, the conditional probability law governing its future depends only on its present state. This property is the probabilistic generalization of the classical notion that, if the present state of a physical system is described in sufficient detail, the system's future evolution would be determined by the present state, without regard to how the system arrived at that state.

The definitions of "time" and "state" depend on the application at hand and on the demands of mathematical tractability. Otherwise, if such practical considerations are ignored, every stochastic process can be made Markovian by enhancing its state space sufficiently.

The theory of Markov processes is the most extensively developed part of probability theory. It covers, in particular, Poisson processes, Brownian motions, and all other Lévy processes. Our aim is to introduce the basic concepts and illustrate them with a few examples and counter-examples. No attempt is made at completeness.

Section 1 is on the Markov property in general. There are examples of Markov chains (discrete-time), of Markov processes (continuous-time), and of anomalous processes lacking the strong Markov property.

Sections 2 and 3 are on two important classes of processes: Itô diffusions and jump-diffusions. They are introduced as solutions to stochastic integral equations. Markov and strong Markov properties are proved directly, generators and resolvents are calculated, and forward and backward equations of Kolmogorov are derived. A quick introduction to stochastic differential equations is given as an appendix in Section 7 for the needs of these sections. These sections can be omitted if the interest is on the general theory.

Markov processes are re-introduced in Section 4 within a modern axiomatic setting. Their Markov property is discussed once more, Blumenthal's zero-one law is proved, the states are classified as holding versus instantaneous, and the behavior at holding states is clarified.

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_9,

Section 5 continues the axiomatic treatment by introducing Hunt processes and Feller processes. The meaning of quasi-left-continuity is explained, the total unpredictability of jump times is given, and the effects of strong Markov property are illustrated.

Section 6 is on resolvents and excessive functions, the connections between them, and their relationships to martingales. It is almost independent of the earlier sections and can be read after Section 2 if desired.

1 Markov Property

Throughout this section, $\mathbb{T}$ is a subset of $\mathbb{R}$; its elements are called times; it will mostly be $\mathbb{R}_+$ and sometimes $\mathbb{N}$. Throughout, $(\Omega,\ \mathcal{H},\ \mathbb{P})$ is a probability space, and $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is a filtration over it.

Let $X = (X_t)_{t\in\mathbb{T}}$ be a stochastic process with some state space $(E,\ \mathcal{E})$ and adapted to the filtration $\mathcal{F}$. We let $\mathcal{G}^o = (\mathcal{G}^o_t)_{t\in\mathbb{T}}$ be the filtration generated by it and put $\mathcal{G}^t_\infty = \sigma\{X_u :\ u \geq t,\ u \in \mathbb{T}\}$, its future after time t.

1.1 Definition. *The process X is said to be Markovian relative to $\mathcal{F}$ if, for every time t, the past $\mathcal{F}_t$ and the future $\mathcal{G}^t_\infty$ are conditionally independent given the present state X_t.*

If X is Markovian relative to $\mathcal{F}$, then it is such relative to $\mathcal{G}^o$ as well, because $\mathcal{G}^o_t \subset \mathcal{F}_t$ by the adaptedness of X to $\mathcal{F}$. It is said to be Markovian, without mentioning a filtration, if it is such relative to $\mathcal{G}^o$.

A similar notion, the strong Markov property, is said to hold if the fixed times t in the preceding definition can be replaced by stopping times. Most Markovian processes are strong Markov, but there are exceptions (see the Examples 1.28 and 1.29).

Characterization

The next theorem uses the definition of conditional independence and properties of repeated conditioning. We use the common shorthand for conditional expectations.

1.2 Theorem. *The following are equivalent:*

a) The process X is Markovian relative to $\mathcal{F}$.

b) For every time t and time $u > t$ and function f in $\mathcal{E}_+$,

1.3 $$\mathbb{E}\,(f \circ X_u \,|\, \mathcal{F}_t) = \mathbb{E}\,(f \circ X_u \,|\, X_t)\,.$$

c) For every time t and positive variable V in $\mathcal{G}^t_\infty$,

1.4 $$\mathbb{E}\,(V|\mathcal{F}_t) = \mathbb{E}\,(V|X_t)\,.$$

d) For every time t and positive variable V in $\mathcal{G}^t_\infty$,

1.5 $$\mathbb{E}\,(V|\mathcal{F}_t) \in \sigma X_t.$$

1.6 REMARK. i) The statement (d) is the closest to the intuitive meaning of the Markov property: estimate of a variable determined by the future is a deterministic function of the present state only (regardless of all the past information) – recall that σX_t is the σ-algebra generated by X_t.

ii) The collection of all f for which 1.3 holds is a monotone class. Thus, the theorem remains true, when, in the statement (b), the condition 1.3 holds for every f in $\mathcal{E}_b$ (bounded $\mathcal{E}$-measurable), or every f in $\mathcal{E}_{b+}$, or every indicator $f = 1_A$ with A in $\mathcal{E}$, or every indicator $f = 1_A$ with A in some p-system generating $\mathcal{E}$.

iii) Similarly, by monotone class arguments again, the theorem remains true if 1.4 (or, equivalently, 1.5) is required only for V having the form

1.7 $$V_n = f_1 \circ X_{u_1} \cdots f_n \circ X_{u_n}$$

with some integer $n \geq 1$, some times $t \leq u_1 < u_2 < \ldots < u_n$, and some functions $f_1, \ldots, f_n$ in $\mathcal{E}_+$. Moreover, the functions f_i can be restricted further as in the preceding remark.

Proof. (a) $\Leftrightarrow$ (c) by the definition of conditional independence; (c) $\Rightarrow$ (b) trivially; and we shall show that (b) $\Rightarrow$ (d) $\Rightarrow$ (c). The last implication is easy: assuming 1.5,

$$\mathbb{E}\,(V|\mathcal{F}_t) = \mathbb{E}\,(\mathbb{E}\,(V|\mathcal{F}_t)\,|X_t) = \mathbb{E}\,(V|X_t)\,.$$

To prove that (b) $\Rightarrow$ (d), assume (b). By Remark 1.6 iii, it is enough to show 1.5 for V having the form 1.7. We do this by induction. For $n = 1$, we have 1.5 from (b). Assume that 1.5 holds for every V_n having the form 1.7. Note that

$$\mathbb{E}\,(V_{n+1}|\mathcal{F}_{u_n}) = V_n\,\mathbb{E}\left(f_{n+1} \circ X_{u_{n+1}}|\mathcal{F}_{u_n}\right) = V_n \cdot g \circ X_{u_n} = \hat{V}_n$$

for some g in $\mathcal{E}_+$ in view of (b). Since $\hat{V}_n$ has the form 1.7 with gf_n replacing f_n, the induction hypothesis applies to $\hat{V}_n$ to yield

$$\mathbb{E}\,(V_{n+1}|\mathcal{F}_t) = \mathbb{E}\left(\hat{V}_n|\mathcal{F}_t\right) \in \sigma X_t.$$

Thus, 1.5 holds for V_{n+1} as well. □

Transition functions

Recall that a Markov kernel on $(E, \mathcal{E})$ is a transition probability kernel from $(E, \mathcal{E})$ into $(E, \mathcal{E})$; see I.6.5 *et seq.* Let $(P_{t,u})$ be a family of Markov

kernels on $(E, \mathcal{E})$ indexed by pairs of times $t \le u$. It is said to be a *Markovian transition function* on $(E, \mathcal{E})$ if

1.8 $$P_{s,t}\ P_{t,u} = P_{s,u}, \qquad 0 \le s < t \le u.$$

The preceding is called the *Chapman-Kolmogorov equation.*

The Markovian process X is said to *admit* $(P_{t,u})$ as a transition function if

1.9 $$\mathbb{E}\,(f \circ X_u | X_t) = (P_{t,u} f) \circ X_t, \qquad t < u, f \in \mathcal{E}_+.$$

Obviously, it is sufficient to check 1.9 for f that are indicators. This provides the intuitive meaning for the kernels:

1.10 $$P_{t,u}\,(x, A) = \mathbb{P}\,\{X_u \in A | X_t = x\}\,.$$

1.11 Remark. There are Markov processes that have no transition functions. Here is an example. Suppose that $\mathbb{T} = \mathbb{R}_+$, and $E = \mathbb{R}_+ \times \Omega$, and $\mathcal{E}$ consists of subsets A of E such that the section $\{\omega \in \Omega : (t, \omega) \in A\}$ belongs to $\mathcal{F}_t$ for every t. Suppose, further, that $X_t(\omega) = (t, \omega)$ for t in $\mathbb{R}_+$ and ω in Ω. Then, $X = (X_t)_{t \in \mathbb{R}_+}$ is a stochastic process with state space $(E, \mathcal{E})$ and is adapted to $\mathcal{F}$. Note that the σ-algebra generated by X_t is exactly $\mathcal{F}_t$, and, hence, the condition 1.3 holds automatically. This Markovian process has no transition function. It is also devoid of interest, since there is nothing further to be said about it.

1.12 Remark. The preceding example illustrates that every process can be made Markovian, but at the cost of mathematical tractability. Begin with a process X^0 with some state space $(D, \mathcal{D})$. Let $\mathcal{F}$ be the filtration generated by it. Define $(E, \mathcal{E})$ and X as in the preceding remark. Now, the "state" of X at time t is the whole history of X^0 until t. By this device, X^0 is converted to the Markovian process X.

Time-homogeneity

Suppose that X is Markovian and admits $(P_{t,u})$ as its transition function. It is said to be *time-homogeneous* if, for every time t and time $u > t$, the dependence of $P_{t,u}$ on the pair (t,u) is through $u - t$ only, that is, if

1.13 $$P_{t,u} = P_{u-t}$$

for some Markov kernel P_{u-t}. Theoretically, there is no loss of generality in assuming time-homogeneity: if X is not, then it can be studied through $\hat{X} = (t, X_t)$, and $\hat{X}$ is Markovian and time-homogeneous. Note that this trick makes time a part of the state description. See Exercise 1.40.

Chains and Processes

Suppose that X is Markovian and time-homogeneous. We call it a *Markov chain* if $\mathbb{T} = \mathbb{N}$, and *Markov process* if $\mathbb{T} = \mathbb{R}_+$.

Suppose that X is a Markov chain. Then, $Q = P_{t,t+1}$ is free of t, and the Chapman-Kolmogorov equation 1.8 yields

1.14 $$P_{t,u} = Q^n, \qquad t \in \mathbb{N},\ u - t = n \in \mathbb{N}.$$

This is expressed by saying that X is a Markov chain with *transition kernel* Q.

Suppose that X is a Markov process. Then, the Markov kernels $P_t,\ t \in \mathbb{R}_+$, must satisfy the semigroup property

1.15 $$P_t\ P_u = P_{t+u}, \qquad t, u \in \mathbb{R}_+,$$

this being the Chapman-Kolmogorov equation in view of 1.13. Then, it is usual to call (P_t) a *transition semigroup* and to say that X is a Markov process with *transition function* (P_t).

For a chain, since the time-set has only one limit point, the analysis required is more straight forward and has more to do with limits in distribution of X_n as $n \to \infty$. For a process, the mathematical treatment has greater ties to classical analysis and semigroups and partial differential equations. We shall concentrate on processes; an analogous program for chains can be carried out without difficulty. However, as a way of displaying the Markov property in its most direct form, we give examples of chains next.

Markov chains

Every Markov chain encountered in applications is constructed from a sequence of independent and identically distributed random variables through a deterministic transformation. In fact, if the state space $(E, \mathcal{E})$ is standard, we may construct every Markov chain in this fashion; see Exercise 1.38 for an illustration with $E = \mathbb{R}$. Interestingly, this form shows that every Markov chain (and, by extension, every Markov process) is a Lévy chain (or Lévy process) in an abstract sense.

Let $(E, \mathcal{E})$ and $(D, \mathcal{D})$ be measurable spaces. Let $\varphi :\ E \times D \mapsto E$ be measurable with respect to $\mathcal{E} \otimes \mathcal{D}$ and $\mathcal{E}$. Let X_0 be a random variable taking values in $(E, \mathcal{E})$ and, independent of it, let $(Z_n)_{n \in \mathbb{N}}$ be an independency of identically distributed variables taking values in $(D, \mathcal{D})$. Define

1.16 $$X_{n+1} = \varphi\left(X_n, Z_{n+1}\right), \qquad n \in \mathbb{N}.$$

Together with X_0, this defines a Markov chain $X = (X_n)_{n \in \mathbb{N}}$ with state space $(E, \mathcal{E})$ and transition kernel

1.17 $$Q(x, A) = \mathbb{P}\left\{\varphi(x, Z_0) \in A\right\}, \qquad x \in E,\ A \in \mathcal{E}.$$

The formula 1.16 encapsulates the essence of Markov chains: the next state X_{n+1} is a deterministic function of the present state X_n and the next random influence Z_{n+1}. The deterministic function φ remains the same over all time;

this is time-homogeneity. In this context, φ is called the structure function and the Z_n are the driving variables. Here are some examples and implications of this construction.

1.18 *Random walks.* Suppose that $E = D = \mathbb{R}^d$ with the attendant Borel σ-algebras $\mathcal{E}$ and $\mathcal{D}$. Take $\varphi(x,\ z) = x + z$. The resulting Markov chain is called a random walk on $\mathbb{R}^d$.

1.19 *Gauss-Markov chains.* Let $E = D = \mathbb{R}^d$ again. Suppose that the Z_n have the d-dimensional standard Gaussian distribution. Take

$$\varphi(x, z) = Ax + Bz, \qquad x, z \in \mathbb{R}^d,$$

where A and B are some $d \times d$ matrices. The resulting chain X is called a Gauss–Markov chain. If X_0 is fixed or has some Gaussian distribution, then the chain (X_n) is Gaussian; this can be seen by noting that

$$X_n = A^n X_o + A^{n-1}B\ Z_1 + \cdots + AB\ Z_{n-1} + B\ Z_n.$$

1.20 *Products of random matrices.* Suppose that $E = D = \mathbb{R}^{d \times d}$, the space of $d \times d$ matrices with real entries. Then, the Z_n are independent and identically distributed random matrices. Take $\varphi(x, z) = zx$, the matrix x multiplied on the left by the matrix z. The resulting chain is given by $X_n = Z_n \cdots Z_1 X_0$; it is a "left random walk" on the set of $d \times d$ matrices. Similarly, taking $\varphi(x, z) = xz$ yields a "right random walk".

1.21 *Continuation.* Suppose that $E = \mathbb{R}^d$ and $D = \mathbb{R}^{d \times d}$; and take $\varphi(x,\ z) = zx$, the product of the matrix z and the column vector x, Then, the chain (X_n) becomes the orbit of the random point X_0 under successive applications of the random linear transformations represented by the matrices $Z_1,\ Z_2, \ldots$.

1.22 *Random dynamical systems.* This is to give a different interpretation to 1.16. Leave $(E, \mathcal{E})$ arbitrary. Define, for each n, a random transformation Φ_n from E into E by letting

$$\Phi_n^\omega(x) = \varphi\left(x, Z_n(\omega)\right), \qquad \omega \in \Omega, x \in E.$$

Then, $\Phi_1, \Phi_2, \ldots$ are independent and identically distributed random transformations from $(E, \mathcal{E})$ into $(E, \mathcal{E})$, and $X_{n+1} = \Phi_{n+1}\ (X_n)$. So, the chain X is obtained by successive applications of independent and identically distributed random transformations.

1.23 *Continuation.* Let $\varphi_{m,n}^\omega$ be the composition of the transformations $\Phi_{m+1}^\omega, \ldots, \Phi_n^\omega$, that is, define

$$\varphi_{m,n}^\omega = \begin{cases} \text{identity} & \text{if } m = n, \\ (\Phi_n^\omega) \circ \cdots \circ \left(\Phi_{m+1}^\omega\right) & \text{if } m < n, \end{cases}$$

for $0 \le m \le n < \infty$. For each ω, the family $\left\{\varphi_{m,n}^\omega : 0 \le m \le n < \infty\right\}$ is a flow, that is, the flow equation

$$\varphi_{m,n}^\omega\left(\varphi_{k,m}^\omega(x)\right) = \varphi_{k,n}^\omega(x), \qquad 0 \le k \le m \le n,$$

is satisfied. And, the Markov chain X is the path of X_0 under the action of the random flow $\varphi = (\varphi_{m,n})_{0 \le m \le n < \infty}$.

Regarding $\varphi_{m,n}$ as the "increment" of φ over the interval $(m,\ n]$, we see that φ has stationary and independent increments. Thus, the Markov chain X is, in this abstract sense, a discrete-time Lévy process in the space of transformations.

Examples of Markov processes

Brownian motion is a Markov process. A number of Markov processes related to it were given in Chapter VIII on Brownian motion; see Examples VIII.1.19, VIII.1.21, VIII.1.22. The following examples are to forge some connections, and also give some pathological (see 1.28) and fascinating (see 1.29) cases where the strong Markov property fails.

1.24 *Lévy processes.* Suppose that $E = \mathbb{R}^d$ and $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$, and assume that $X_t = X_0 + Y_t,\ t \in \mathbb{R}_+$, where $Y = (Y_t)$ is a Lévy process independent of X_0. Let π_t be the distribution of Y_t, and recall that $A - x = \{y - x :\ y \in A\}$. Then, X is a Markov process with transition function

1.25 $$P_t(x, A) = \pi_t(A - x), \qquad x \in E,\ A \in \mathcal{E},\ t \in \mathbb{R}_+.$$

In other words, X is both time-homogeneous and spatially homogeneous. Conversely, if X is such, that is, if X is a Markov process whose transition semigroup (P_t) has the form 1.25, then $X = X_0 + Y$ for some Lévy process Y.

1.26 *Markov chains subordinated to Poisson.* Let $(Y_n)_{n\in\mathbb{N}}$ be a Markov chain with state space $(E,\ \mathcal{E})$ and transition kernel Q. Let (N_t) be a Poisson process, with rate c, independent of the chain (Y_n). Suppose that

$$X_t = Y_{N_t}, \qquad t \in \mathbb{R}_+.$$

Then, X is a Markov process with state space (E, $\mathcal{E}$) and transition function (P_t), where

1.27 $$P_t(x, A) = \sum_{n=0}^{\infty} \frac{e^{-ct}(ct)^n}{n!} Q^n(x, A).$$

1.28 *Delayed uniform motion.* The state space is $E = \mathbb{R}_+$. The process depicts the motion of a particle that is at the origin initially, stays there an exponentially distributed amount T of time, and then moves upward at unit speed:

$$X_t = (t - T)^+, \qquad t \in \mathbb{R}_+.$$

This X is a Markov process. Its transition function (P_t) is easy to compute by using the working formula

$$P_t f(x) = \mathbb{E}\left(f \circ X_{s+t} | X_s = x\right):$$

If $x > 0$, then $X_{s+t} = x + t$. If $x = 0$, the particle's sojourn at 0 has not ended yet, that is, $T > s$. By the exponential nature of T, then, the remaining sojourn time $T - s$ has the same exponential distribution as T itself. Letting c be the parameter of that exponential distribution, we get

$$P_t f(x) = \begin{cases} f(x+t) & if\ x > 0, \\ e^{-ct} f(0) + \int_0^t du\ ce^{-cu}\ f(t-u) & if\ x = 0. \end{cases}$$

Suppose now that the filtration $\mathcal{F}$ is taken to be $\left(\mathcal{G}^o_{t+}\right)$. Then, T is a stopping time of $\mathcal{F}$, and $X_T = X_0 = 0$. If X were strong Markov, the future after T would have the same law as the future at $t = 0$. But it is not so; future at $t = 0$ starts with a sojourn of some length at 0, whereas the future at T is that of immediate motion. So, this process is *not* strong Markov.

Intuitive notion of the Markov property is that the present state determines the law of the future; and this is tacitly extended to cases where "the present time" is allowed to be a stopping time. The present example is cautionary. At the same time, it displays the reason for the failure of the strong Markov property: the state 0 is allowed to play two different roles: as a point of sojourn, and as a launching pad for the motion. If we re-define the process as

$$\hat{X}_t(\omega) = \begin{cases} -1 & \text{if } t < T(\omega), \\ t - T(\omega) & \text{if } t \geq T(\omega), \end{cases}$$

then we have a strong Markov process $\hat{X}$ with state space $\{-1\} \cup \mathbb{R}_+$.

1.29 *Lévy's increasing continuous process.* This is an example, due to Lévy, of another process whose Markov property does not extend to stopping times. Moreover, it illustrates the importance of choosing the correct state space and the correct construction of the process. As a by-product, it shows the advantages of concentrating on the dynamics of the random motion, instead of the analytic machinery of transition functions and the like.

The canonical process has state space $\mathbb{R}_+$. Started at 0, its paths are increasing continuous with limit $+\infty$ as time goes to $+\infty$. Every rational number in $\mathbb{R}_+$ is a holding point, that is, the process has an exponential sojourn there before resuming its upward creep. The process spends no time in the set of irrationals. By re-labeling the states, we shall get a bizarre Markov process with state space $\bar{\mathbb{N}} = \{0, 1, \ldots, +\infty\}$.

Let $\mathbb{Q}_+$ denote the set of rational numbers in $\mathbb{R}_+$. Over the probability space $(\Omega, \mathcal{H}, \mathbb{P})$, we suppose that $\{Z_q : q \in \mathbb{Q}_+\}$ is an independency of $\mathbb{R}_+$-valued exponential random variables with Z_q having the mean $m(q)$, where

$$\sum_{q \in \mathbb{Q}_+ \cap [0,1)} m(q) = 1,$$

and, for each integer $n \geq 1$, we have $m(q) = m(q - n)$ for q in $[n, n+1)$. We are thinking of a particle that moves upward in $\mathbb{R}_+$, having a sojourn

of length Z_q at each rational q, and spending no time elsewhere. Thus, the cumulative time it spends in the set $[0,x]$ is

1.30 $$S_x = \sum_{q \in \mathbb{Q}_+ \cap [0,\, x]} Z_q, \qquad x \in \mathbb{R}_+,$$

and, therefore, the particle's position at time t is

1.31 $$X_t = \inf \{x \in \mathbb{R}_+ : S_x > t\}, \qquad t \in \mathbb{R}_+.$$

In view of the way the means $m(q)$ are chosen, for almost every ω, we have $S_x(\omega) < \infty$ for all x in $\mathbb{R}_+$, but with limit $+\infty$ as $x \to \infty$. Clearly, $x \mapsto S_x(\omega)$ is right-continuous and strictly increasing, which implies that $t \mapsto X_t(\omega)$ is continuous increasing. Moreover, since the path $S(\omega)$ is of the pure-jump type,

1.32 $$\text{Leb}\,\{t \in \mathbb{R}_+ : X_t(\omega) \notin \mathbb{Q}_+\} = 0.$$

The process $(S_x)_{x \in \mathbb{R}_+}$ is a pure-jump process with independent (but non-stationary) increments; see Exercise VI.4.24. It jumps at every rational q by the exponential amount Z_q. Thus, $X = (X_t)_{t \in \mathbb{R}_+}$ is a Markov process (time-homogeneous) with a transition function (P_t) that can be specified, see Exercise 1.39; it is Markov relative to $\left(\mathcal{G}^o_{t+}\right)$ as well.

Heuristically, given that $X_t = x$ and $x \in \mathbb{Q}_+$, then, the particle will stay at x a further amount of time that is exponential with mean $m(x)$ and then start its upward motion. This is for fixed time t. But, when t is replaced by the random time T at which the particle departs the fixed rational point x, the future looks different. Thus, X lacks the strong Markov property.

Next, we take advantage of 1.32 to define a Markov process with a discrete state space. Let b be a bijection from $\mathbb{Q}_+$ onto $\mathbb{N}$; this is just a re-labeling of the rationals by integers. Define

1.33 $$Y_t(\omega) = \begin{cases} b \circ X_t(\omega) & if\ X_t(\omega) \in \mathbb{Q}_+, \\ +\infty & \text{otherwise.} \end{cases}$$

Then, $Y = (Y_t)$ is a Markov process with state space $\bar{\mathbb{N}}$. Its paths are difficult to describe directly: if $Y_t = i$, then the particle stays there an exponential time, but there is no "next" integer state to go. The state $+\infty$ is "fictitious"; the total amount of time spent there by Y is zero by 1.32. The paths have discontinuities of the second kind. We may define

$$\mathrm{Q}_t(i, A) = \mathbb{P}\{Y_{s+t} \in A | Y_s = i\}, \qquad t \in \mathbb{R}_+, i \in \mathbb{N},\ A \subset \mathbb{N},$$

to obtain a Markov transition semigroup, that is, $Q_t Q_u = Q_{t+u}$ and $Q_t(i, \mathbb{N}) = 1$ for each i. For this reason, Y is said to have $\mathbb{N}$ as its *minimal state space*. This process is a good example of inadequacy of transition functions (and generators to come) as the base to build the theory on. Despite this sentiment, we continue with ...

Probability Laws

We return to the general case with index set $\mathbb{T}$ and suppose that X is Markovian and admits $(P_{t,u})$ as its transition function. Suppose, further, that $\mathbb{T} \subset \mathbb{R}_+$ and $0 \in \mathbb{T}$. Let μ_0 be the distribution of X_0. Then, for times $0 = t_0 < t_1 < \cdots < t_n$,

1.34
$$\begin{aligned}&\mathbb{P}\left\{X_{t_0} \in dx_0, X_{t_1} \in dx_1, X_{t_2} \in dx_2, \ldots, X_{t_n} \in dx_n\right\} \\ &= \mu_{t_0}(dx_0) P_{t_0,t_1}(x_0, dx_1) P_{t_1,t_2}(x_1, dx_2) \cdots P_{t_{n-1},t_n}(x_{n-1}, dx_n)\,.\end{aligned}$$

This follows from repeated applications of the Markov property. It shows, as well, that the probability law of X is determined by the initial distribution μ_0 and the transition function $(P_{t,u})$. Modern theory treats $(P_{t,u})$ as fixed, but μ_0 as a variable; it is usual to write $\mathbb{P}^\mu$ for $\mathbb{P}$ when $\mu_0 = \mu$, and $\mathbb{P}^x$ when $\mu_0 = \delta_x$, Dirac at x.

Existence and construction

Let μ be a probability measure and $(P_{t,u})$ a Markov transition function, both on some measurable space $(E, \mathcal{E})$. If $\mathbb{T} = \mathbb{N}$, then Theorem IV.4.7 shows the existence of a probability space $(\Omega, \mathcal{H}, \mathbb{P}^\mu)$ and a process $X = (X_t)_{t \in \mathbb{T}}$ such that X is Markovian with initial distribution μ and transition function $(P_{t,u})$, that is, such that 1.34 holds. If $\mathbb{T} = \mathbb{R}_+$, the same existence result follows from the Kolmogorov extension theorem, IV.4.18, under a slight condition on $(E, \mathcal{E})$. We refer to Chapter IV, Sections 4 and 5, for the details as well as for a discussion of some special cases and issues regarding "time" and "space".

In practice, however, one rarely has $(P_{t,u})$ specified from the start. Instead, X is constructed from well-known objects, and $(P_{t,u})$ is defined implicitly from X. For instance, the example 1.28 is constructed from one exponential variable, and the example 1.29 from a countable independency of exponentials. As we know, Wiener processes and Poisson random measures on $\mathbb{R}^2$ can be constructed from a countable independency of uniform variables, and Lévy processes are constructed from Wiener processes and Poisson random measures. Similarly, most Markov processes are constructed from a countable independency of uniform variables via Wiener processes and Poisson random measures; the constructions are sometimes direct, and often by means of stochastic integral equations. Sections 2 and 3 illustrate the method.

Exercises and complements

1.35 *Processes with discrete state spaces.* Suppose that X is a Markov process (time-set $\mathbb{R}_+$, time-homogeneous) with state space $(E, \mathcal{E})$ and transition

function (P_t). Suppose that $(E, \mathcal{E})$ is discrete, that is, E is countable and $\mathcal{E} = 2^E$, the discrete σ-algebra on E. Then, each P_t has the form

$$P_t(x, A) = \sum_{y \in A} p_t(x, y), \qquad x \in E,\ A \in \mathcal{E},$$

and we may regard $y \mapsto p_t(x, y)$ as the density of the measure $A \mapsto P_t(x, A)$ with respect to the counting measure on $(E, \mathcal{E})$. Of course, then, we may identify the kernel P_t with the matrix whose entries are the probabilities $p_t(x, y)$. We shall do this without further comment.

1.36 *Continuation.* Suppose that E consists of two elements, a and b. Let

$$P_t = \begin{bmatrix} q + pe^{-ct} & p - pe^{-ct} \\ q - qe^{-ct} & p + qe^{-ct} \end{bmatrix}, \qquad t \geq 0,$$

where p and q are numbers in [0,1] with $p + q = 1$, and c is a number in $\mathbb{R}_+$. Show that the matrices P_t satisfy $P_0 = I$ and $P_t P_u = P_{t+u}$. When E consists of two states, this is the most general form of a transition function (P_t). The case $c = 0$ is degenerate (what happens then?). Describe the paths in the cases $p = 0$ or $q = 0$.

1.37 *Subordination of Markov to Lévy.* Let X be a Markov process with state space $(E, \mathcal{E})$ and transition function (P_t). Let $S = (S_t)$ be an increasing Lévy process independent of X, and with distribution π_t for S_t. Define

$$\hat{X}_t = X_{S_t}, \qquad t \in \mathbb{R}_+.$$

Show that $\hat{X}$ is again a Markov process with state space $(E, \mathcal{E})$. Compute its transition function $(\hat{P}_t)$ in terms of (P_t) and (π_t).

1.38 *Markov chains.* Let Q be a Markov kernel on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. For each x in $\mathbb{R}$, define

$$\varphi(x, u) = \inf \{y \in \mathbb{R} : Q(x, (-\infty, y]) > u\}, \qquad u \in (0, 1).$$

Then, $u \mapsto \varphi(x, u)$ is increasing and right-continuous.

a) Show that $x \mapsto \varphi(x, u)$ is Borel measurable for each u. Conclude that φ is a Borel function on $\mathbb{R} \times (0, 1)$.

b) Let (Z_n) be an independency of uniform variables taking values in (0,1). Suppose X_0 is independent of (Z_n), and define (X_n) by 1.16. Show that (X_n) is a Markov chain with transition kernel Q.

1.39 *Lévy's example.* Let X be as in Example 1.29. Let (P_t) be its transition function. Show that, for real numbers $0 \leq x \leq y$,

$$P_t(x, (y, \infty)) = \mathbb{P}\left\{ \sum_{x \leq q \leq y} Z_q < t \right\}$$

where the sum is over the rationals q in the interval $[x, y]$. Show that

$$\int_0^\infty dt\ e^{-pt} P_t\left(x, (y, \infty)\right) = \frac{1}{p} \prod_{x \le q \le y} \frac{1}{1 + m(q)p}, \quad p \in \mathbb{R}_+.$$

This specifies (P_t), at least in principle.

1.40 *Time-homogeneity.* Suppose that X is Markovian with state space $(E, \mathcal{E})$ and admits $(P_{t,u})$ as its transition function (we do not assume time-homogeneity). Define

$$\hat{X}_t = (t, X_t)\,, \qquad t \in \mathbb{R}_+.$$

Then, $\hat{X}$ is Markovian with state space $(\hat{E}, \hat{\mathcal{E}}) = (\mathbb{R}_+ \times E,\ \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{E})$. Show that it is time-homogeneous. Show that its transition function $(\hat{P}_t)$ is given by, for positive f in $\hat{\mathcal{E}}$,

$$\hat{P}_t f\left(\hat{x}\right) = \int_E P_{s,s+t}(x, dy) f(s+t, y), \qquad \hat{x} = (s, x) \in \hat{E}.$$

1.41 *Processes with independent increments.* Let X be a process in $\mathbb{R}^d$ having independent increments, but without stationarity of increments. Then, X is Markovian, but without time-homogeneity; we have

$$P_{t,u}(x, A) = \mathbb{P}\left\{X_u \in A | X_t = x\right\} = \mathbb{P}\left\{X_u - X_t \in A - x\right\} = \pi_{t,u}\left(A - x\right).$$

Define $\hat{X}_t = (t,\ X_t)$ as in the preceding example. Then, $\hat{X}$ is a time-homogeneous Markov process. Compute its transition function $(\hat{P}_t)$ in terms of $\pi_{s,t}$. Note that $\hat{X}$ has independent increments, but still without the stationarity of increments.

1.42 *Expanded filtrations.* Suppose that X is Markovian relative to the filtration $\mathcal{F}$. Let $\mathcal{H}_0$ be a sub-σ-algebra of $\mathcal{H}$ that is independent of $\mathcal{F}_\infty$. Put $\hat{\mathcal{F}}_t = \mathcal{H}_0 \vee \mathcal{F}_t$ for each time t. Then, X is Markovian relative to $\hat{\mathcal{F}}$ as well. Show. This is helpful when X is being studied in the presence of other processes that are independent of X.

1.43 *Entrance laws.* Suppose that X is Markovian with time-set $\mathbb{T}$ and transition function $(P_{t,u})$. Suppose that $\mathbb{T}$ does not have an initial element; $\mathbb{T} = (0, \infty)$ or $\mathbb{T} = (-\infty, +\infty)$ for instance. Let μ_t be the distribution of X_t. Then, the formula 1.34 holds for times $t_0 < t_1 < \ldots < t_n$ in $\mathbb{T}$. Note that, necessarily,

1.44 $$\int_E \mu_t(dx) P_{t,u}\left(x, A\right) = \mu_u\left(A\right), \qquad t < u, A \in \mathcal{E}.$$

In general, if a family (μ_t) of probability measures satisfies 1.44 for some transition function $(P_{t,u})$, then (μ_t) is said to be an *entrance law* for $(P_{t,u})$. If $\mathbb{T}$ has an initial element t_0, then $\mu_t = \mu_{t_0}\ P_{t_0,t},\ t \ge t_0$, defines an entrance law (μ_t) from the initial law μ_{t_0}.

2 Itô Diffusions

Itô diffusions are continuous strong Markov processes satisfying certain stochastic differential equations. They are generalizations of Brownian motions in the following way.

Over some probability space, let X be a Brownian motion on $\mathbb{R}$. It has the form $X_t = X_0 + at + b\, W_t$, where W is a Wiener process, and a and b constants. The dynamics of the motion is expressed better in the classical fashion:

$$dX_t = a\ dt + b\ dW_t,$$

that is, velocity is equal to a constant a perturbed by some "noise." We notice that X will remain Markovian in the more general case where a is replaced with $a(X_t)$, some function of the current position X_t, and the noise multiplier b is replaced with $b(X_t)$, some function of X_t. The result is

2.1
$$dX_t = a \circ X_t\ dt + b \circ X_t\ dW_t$$

or, equivalently, in the formal language of integrals,

2.2
$$X_t = X_0 + \int_0^t a \circ X_s\ ds + \int_0^t b \circ X_s\ dW_s.$$

But there arises a problem: the second integral does not have a conventional meaning, because the paths $t \mapsto W_t$ have infinite total variation over every interval $[s,\ s+u]$ with $u > 0$. Fortunately, it is possible to give a meaning to such integrals, called stochastic integrals of Itô to distinguish them from the ordinary ones.

This section can be read without previous exposure to stochastic calculus if one is willing to take some results on faith. Nevertheless, we put a summary of stochastic integration, as an appendix, in Section 7.

Stochastic base

The motion of interest will be a continuous process with state space $(E,\ \mathcal{E})$, where $E = \mathbb{R}^d$ for some fixed dimension $d \geq 1$ and $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$. The process will be the solution of a stochastic differential equation driven by a multi-dimensional Wiener process.

Throughout this section, $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space, $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is an augmented right-continuous filtration, and $W = (W^1, \ldots, W^m)$ is an m-dimensional Wiener process adapted to $\mathcal{F}$; the integer m will remain fixed. In addition, X_0 is an E-valued random variable in $\mathcal{F}_0$ and is, thus, independent of W. We let $(x, H) \mapsto \mathbb{P}^x(H)$ be a regular version of the conditional probabilities

2.3
$$\mathbb{P}^x(H) = \mathbb{P}\left(H | X_0 = x\right).$$

Equation of motion

The deterministic data are some vector fields $u_0, \ldots, u_m$ on $E = \mathbb{R}^d$, that is, each u_n is a mapping from E into E. Throughout, we assume that the following condition of *Lipschitz continuity* holds; here $|x|$ is the length of the vector x for each x in E.

2.4 Condition. There is a constant c in $\mathbb{R}_+$ such that

$$|u_n(x) - u_n(y)| \leq c\,|x - y|\,, \qquad x, y \in E,\ 0 \leq n \leq m. \qquad \square$$

This condition ensures that the following equation of motion makes sense and has a unique solution (see Theorem 2.13 below):

$$2.5 \qquad X_t = X_0 + \int_0^t u_0 \circ X_s\ ds + \sum_{n=1}^m \int_0^t u_n \circ X_s \quad dW_s^n;$$

here, the integrals involving the W^n are to be understood as Itô integrals of stochastic calculus (see Section 7, Appendix).

Somewhat more explicitly, writing X_t^i for the i-component of X_t, and $u_n^i(x)$ for the i-component of the vector $u_n(x)$, the stochastic integral equation 2.5 becomes

$$2.6 \qquad X_t^i = X_0^i + \int_0^t u_0^i \circ X_s\ ds + \sum_{n=1}^m \int_0^t u_n^i \circ X_s\ dW_s^n, \quad 1 \leq i \leq d.$$

Again equivalently, 2.5 can be written as a stochastic differential equation:

$$2.7 \qquad dX_t = u_0 \circ X_t\ dt + \sum_{n=1}^m u_n \circ X_t\ dW_t^n.$$

The looks of the preceding line can be simplified: put $a(x) = u_0(x)$ and let $b(x)$ be the $d \times m$ matrix whose (i,n)-entry is $u_n^i(x)$; then 2.7 becomes

$$2.8 \qquad dX_t = a \circ X_t\ dt + b \circ X_t\ dW_t,$$

which looks exactly like 2.1, and 2.5 gets to look like 2.2. But, 2.5 and 2.7 are better at conveying the role of each W^n: the effect of W^n is carried to the motion by the vector field u_n; this issue becomes important when we consider a cloud of particles whose motions satisfy the same differential equation 2.7.

Examples

2.9 *Geometric Brownian motion.* With $d = m = 1$, and b and c constants in $\mathbb{R}$, consider the geometric Brownian motion

$$X_t = X_0 \exp\left(bW_t + ct\right), \qquad t \in \mathbb{R}_+.$$

Using Itô's formula (Theorem 7.20), we see that X is the solution to

$$dX_t = aX_t\ dt + bX_t\ dW_t,$$

where $a = c + \frac{1}{2}\ b^2$. In particular, when $a = 0$, we obtain the exponential martingale $X_t = X_o \exp(bW_t - \frac{1}{2}b^2 t)$ as the solution to $dX_t = bX_t\ dW_t$.

2.10 *Ornstein–Uhlenbeck process.* In 2.8, suppose that $a(x) = Ax$ and $b(x) = B$, where A and B are matrices of dimensions $d \times d$ and $d \times m$ respectively; we get

$$dX_t = AX_t\ dt + B\ dW_t,$$

which is also called the *Langevin equation.* The solution is

$$X_t = e^{tA} X_0 + \int_0^t e^{(t-s)A} B\ dW_s,$$

where $e^{tA} = \sum_{k=0}^{\infty} \left(t^k/k!\right) A^k$. In the particular case where $d = m = 1$, the matrices reduce to real numbers; and assuming that A is a negative constant, say $A = -c$ and $B = b$, we obtain

$$X_t = e^{-ct} X_0 + b \int_0^t e^{-c(t-s)}\ dW_s, \quad t \in \mathbb{R}_+.$$

This is the one-dimensional velocity process in the model of Ornstein and Uhlenbeck for the physical Brownian motion; see Exercise 2.60 also.

2.11 *Brownian motion on the unit circle.* This is the motion X, on the unit circle in $\mathbb{R}^2$, whose components are

$$X_t^1 = \cos W_t, \quad X_t^2 = \sin W_t,$$

where W is a Wiener process; one can think of it as the complex-valued motion $\exp(iW_t)$. Using Itô's formula, Theorem 7.20, we see that X satisfies (here $d = 2$ and $m = 1$)

$$dX_t = a \circ X_t\ dt + b \circ X_t\ dW_t$$

where $a(x) = -\frac{1}{2}(x_1, x_2)$ and $b(x) = (-x_2, x_1)$ for $x = (x_1, x_2)$.

2.12 *Correlated Brownian motions.* With $d = 1$ and $m = 2$, consider the equation 2.7 with $u_0 = 0,\ u_1 = \sin,\ u_2 = \cos$, that is, consider

$$dX_t = (\sin X_t)\, dW_t^1 + (\cos X_t)\, dW_t^2.$$

This process X is a continuous martingale and its quadratic variation has the differential (see Example 7.5, Theorem 7.15, and Lemma 7.22 for these)

$$(dX_t)^2 = (\sin X_t)^2\ dt + (\cos X_t)^2\ dt = dt.$$

It follows from Theorem 7.24 that $X - X_0$ is a Wiener process. For studying X, then, writing $X = X_0 + \hat{W}$ would be simpler. But this simple description

is inadequate for describing two motions under the same regime. For instance, in addition to X with $X_0 = x$, let Y satisfy the same equation with $Y_0 = y$, that is, with the same W^1 and W^2 as for X,

$$dY_t = (\sin Y_t)\ dW_t^1 + (\cos Y_t)\, dW_t^2.$$

Then, X and Y are standard Brownian motions, but they depend on each other. Their correlation structure is specified by the cross variation process $\langle X, Y\rangle$, which is given by (in differential form)

$$dX_t\ dY_t = (\sin X_t)\,(\sin Y_t)\ dt + (\cos X_t)\,(\cos Y_t)\, dt = \cos\,(X_t - Y_t)\ dt.$$

Existence and uniqueness

Consider the stochastic integral equation 2.5 under Condition 2.4 on the vector fields u_n. As with deterministic differential equations, Lipschitz continuity 2.4 ensures the existence of a unique solution (in the sense to be explained shortly). The method of solution is also the same as in the deterministic case, namely, Pickard's method of successive approximations. The result is listed next; its proof is delayed to 2.52.

2.13 THEOREM. *The equation 2.5 has a pathwise unique solution X; the process X is continuous.*

2.14 REMARK. The proof 2.52 will also show that X is a *strong solution* in the following sense, thus explaining *pathwise uniqueness*: There exists a unique mapping

$$\varphi : E \times C\,(\mathbb{R}_+ \mapsto \mathbb{R}^m) \mapsto C\,(\mathbb{R}_+ \mapsto E)$$

such that, for almost every ω, the paths $X(\omega):\ t \mapsto X_t(\omega)$ and $W(\omega):\ t \mapsto W_t(\omega) = \big(W_t^1(\omega), \dots,\ W_t^m(\omega)\big)$ satisfy

$$X(\omega) = \varphi\,(X_0(\omega), W(\omega))\,.$$

Markov property

The next theorem shows that X is a (time-homogeneous) Markov process with state space E and transition function (P_t), where

2.15 $$P_t f(x) = \mathbb{E}^x\ f \circ X_t, \quad x \in E,\ f \in \mathcal{E}_+,\ t \in \mathbb{R}_+.$$

2.16 THEOREM. *For each t in $\mathbb{R}_+$, the process $\hat{X} = (X_{t+u})_{u\in\mathbb{R}_+}$ is conditionally independent of $\mathcal{F}_t$ given X_t; moreover, given that $X_t = y$, the conditional law of $\hat{X}$ is the same as the law of X under $\mathbb{P}^y$.*

2.17 REMARK. The claim of the theorem is that, for every integer $k \geq 1$ and positive Borel function f on E^k,

$$\mathbb{E}^x\,(f\,(X_{t+u_1}, \dots, X_{t+u_k})\,|\mathcal{F}_t) = \mathbb{E}^{X_t} f\,(X_{u_1}, \dots, X_{u_k})\,,$$

where the right side stands for $g \circ X_t$ with $g(y) = \mathbb{E}^y f(X_{u_1}, \ldots, X_{u_k})$. Of course, as in Theorem 1.2, this is also equivalent to

$$\mathbb{E}^x \left(f \circ X_{t+u} | \mathcal{F}_t \right) = P_u f \circ X_t, \quad x \in E,\ f \in \mathcal{E}_+,\ t, u \in \mathbb{R}_+.$$

Proof. Fix t and let $\hat{W} = (W_{t+u} - W_t)_{u \in \mathbb{R}_+}$. Note that, in the notation system of 2.8,

$$\begin{aligned} \hat{X}_u &= X_t + \int_t^{t+u} a \circ X_s\, ds + \int_t^{t+u} b \circ X_s\, dW_s \\ &= \hat{X}_0 + \int_0^u a \circ \hat{X}_s\, ds + \int_0^u b \circ \hat{X}_s\, d\hat{W}_s. \end{aligned}$$

Thus, with φ defined as in Remark 2.14,

$$X = \varphi\left(X_0, W\right), \qquad \hat{X} = \varphi\left(\hat{X}_0, \hat{W}\right).$$

By the Lévy nature of W, the process $\hat{W}$ is independent of $\mathcal{F}_t$ and is again a Wiener process just as W. Thus, $\hat{X}$ is conditionally independent of $\mathcal{F}_t$ given $\hat{X}_0 = X_t$. Moreover, given that $X_t = \hat{X}_o = y$, the conditional law of $\hat{X}$ is the law of $\varphi(y, \hat{W})$, which is in turn the same as the law of $\varphi(y, W)$, namely, the law of X given $X_0 = y$. □

Strong Markov property

The preceding theorem remains true when the deterministic time t is replaced with a stopping time T, provided that we make provisions for the possibility that T might take the value $+\infty$. To that end we introduce the following.

2.18 Convention. *Let ∂ be a point outside E; put $\bar{E} = E \cup \{\partial\}$, and let $\bar{\mathcal{E}}$ be the σ-algebra on $\bar{E}$ generated by $\mathcal{E}$. We define $X_\infty(\omega) = \partial$ for all ω. Every function $f : E \mapsto \mathbb{R}$ is extended onto $\bar{E}$ by setting $f(\partial) = 0$. If the original f is in $\mathcal{E}_+$, for instance, then the extended function is in $\bar{\mathcal{E}}_+$, but we still write $f \in \mathcal{E}_+$. The convention applies to the function $P_t f$ as well: $P_t f(\partial) = 0$. Finally, $\mathcal{F}_\infty = \lim \mathcal{F}_t = \vee_t \mathcal{F}_t$ as usual.*

2.19 Theorem. *The process X is strong Markov: For every stopping time T of $\mathcal{F}$, the variable X_T is $\mathcal{F}_T$ – measurable, and the process $\hat{X} = (X_{T+u})_{u \in \mathbb{R}_+}$ is conditionally independent of $\mathcal{F}_T$ given X_T; moreover, for y in E, on the event $\{X_T = y\}$, the conditional law of $\hat{X}$ given X_T is the same as the law of X under $\mathbb{P}^y$.*

Proof. Since X is continuous and adapted to $\mathcal{F}$, the random variable X_T is measurable with respect to $\mathcal{F}_T$ and $\bar{\mathcal{E}}$. The rest of the proof follows that of the last theorem: replace t by T throughout to handle the conditional expectations on the event $\{T < \infty\}$. On the event $\{T = \infty\}$, we have $X_{T+u} = \partial$ for all u, and the claim holds trivially in view of the conventions 2.18. □

Generator

We introduce a differential operator which will describe the differential structure of the transition function (P_t). First, some notation: We put

2.20 $\mathcal{C} = C(E\mapsto\mathbb{R}), \mathcal{C}_K = C_K\,(E\mapsto\mathbb{R})\,,\ \mathcal{C}^2 = C^2(E\mapsto\mathbb{R}),\ \mathcal{C}^2_K = \mathcal{C}^2 \cap \mathcal{C}_K;$

Thus, $\mathcal{C}$ is the set of all continuous functions $f:\ E \mapsto \mathbb{R}$, and $\mathcal{C}_K$ is the set of such f with compact support, and $\mathcal{C}^2$ is the set of such f that are twice differentiable with continuous derivatives of first and second order. For f in $\mathcal{C}^2$, we write $\partial_i f$ for the partial derivative with respect to the ith argument, and $\partial_{ij} f$ for the second order partial derivative with respect to the ith and j^{th} arguments; the classical notations are $\partial_i = \frac{\partial}{\partial x_i}$ and $\partial_{ij} = \frac{\partial^2}{\partial x_i \partial x_j}$. When $d = 1$, these become f', the derivative, and f'', the second derivative. With these notations, we introduce the operator G on $\mathcal{C}^2$ by

2.21 $$Gf(x) = \sum_{i=1}^{d} u_0^i(x)\partial_i f(x) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{n=1}^{m} u_n^i(x)u_n^j(x)\,\partial_{ij} f(x), \quad x \in E.$$

When $d = 1$ (and more generally with proper interpretation, in the notation system of 2.8) this becomes

$$Gf(x) = a(x)f'(x) + {}^1\!/\!_2\; b(x)^2\; f''(x)\,.$$

2.22 Example. *Brownian motion.* Suppose that X is a standard Brownian motion in $\mathbb{R}^d$, that is, take $m = d$ and put $u_0(x) = 0$ and let $u_n^i(x)$ be free of x and equal to 1 or 0 according as $i = n$ or $i \neq n$. Then

$$Gf = \frac{1}{2}\sum_{i=1}^{d} \partial_{ii} f, \qquad f \in \mathcal{C}^2;$$

thus, $Gf = {}^1\!/\!_2 \Delta f$, where Δ is the Laplacian operator of classical analysis.

Itô's formula

The equation 2.5 of motion shows that X is a semimartingale. Applying to it Itô's formula, Theorem 7.20, yields the following.

2.23 Theorem. *For every f in $\mathcal{C}^2_K$,*

2.24 $$M_t = f \circ X_t - f \circ X_0 - \int_0^t ds\; Gf \circ X_s, \qquad t \in \mathbb{R}_+,$$

is a martingale; it is given by

$$M_t = \sum_{n=1}^{m}\sum_{i=1}^{d} \int_0^t \left(u_n^i \circ X_s\right)\left(\partial_i f \circ X_s\right)\, dW_s^n.$$

Proof. We use Itô's formula, Theorem 7.20:

$$d\,(f \circ X_t) = \sum_{i=1}^{d} (\partial_i f \circ X_t)\, dX_t^i + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (\partial_{ij} f \circ X_t)\, dX_t^i \, dX_t^j.$$

In view of 2.6 for X_t^i, this yields the claim once we note that

$$dX_t^i \, dX_t^j = \sum_{n=1}^{m} \left(u_n^i \circ X_t\right) \left(u_n^j \circ X_t\right) \, dt$$

in view of the rules of Theorem 7.19 and Lemma 7.22. □

2.25 COROLLARY. *Let $f \in \mathcal{C}_K^2$. Then, $Gf \in \mathcal{C}_K$, and*

$$\mathbb{E}^x \; f \circ X_t = f(x) + \mathbb{E}^x \int_0^t ds \; Gf \circ X_s, \quad x \in E, \; t \in \mathbb{R}_+.$$

Proof. The vector fields u_n are continuous by Condition 2.4. Thus, for f in $\mathcal{C}_K^2$, the formula 2.21 shows that Gf is continuous and has compact support. The claimed formula is now immediate from the preceding theorem, since $\mathbb{E}^x M_t = 0$. □

Moreover, in the preceding corollary, since $Gf \in \mathcal{C}_K$ and thus is bounded, we may change the order of integration and expectation (by Fubini's theorem). Recalling 2.15, then, we obtain the following.

2.26 COROLLARY. *Let $f \in \mathcal{C}_K^2$. Then, $Gf \in \mathcal{C}_K$ and*

$$P_t f(x) = f(x) + \int_0^t ds \; P_s \; Gf(x), \qquad x \in E, \; t \in \mathbb{R}_+.$$

Dynkin's formula

This is essentially Corollary 2.25, but with a stopping time replacing the deterministic time.

2.27 THEOREM. *Let $f \in \mathcal{C}_K^2$. Let T be an $\mathcal{F}$-stopping time. For fixed x in E, suppose that $\mathbb{E}^x T < \infty$; then,*

$$\mathbb{E}^x \; f \circ X_T = f(x) + \mathbb{E}^x \int_0^T ds \; Gf \circ X_s.$$

Proof. Let f, T, x be as described. By Theorem 2.23, the proof is reduced to showing that $\mathbb{E}^x M_T = 0$ for the martingale M there. Since M is a sum of finitely many martingales, it is enough to show that $\mathbb{E}^x \hat{M}_T = 0$ for one of the terms there, say, for the martingale

$$\hat{M}_t = \int_0^t g \circ X_s \; d\hat{W}_s,$$

where, for fixed i and n, we put $g = u_n^i \partial_i f$ and $\hat{W} = W^n$.

Since f has compact support, and since $\partial_i f$ and u_n^i are continuous, the function g is continuous and bounded, say by c. Thus, applying 7.6 with F there taken as the bounded left-continuous process $s \mapsto (g \circ X_s)1_{\{s \le T\}}$,

$$\begin{aligned}\mathbb{E}^x \left(\hat{M}_{T\wedge t}\right)^2 &= E^x \left(\int_0^t g \circ X_s \; 1_{\{s\le T\}} \; d\hat{W}_s\right)^2 \\ &= \mathbb{E}^x \int_0^t (g\circ X_s)^2 \; 1_{\{s\le T\}} \, ds \le \mathbb{E}^x c^2 T < \infty\end{aligned}$$

by the assumption that $\mathbb{E}^x T < \infty$. So, on $(\Omega, \mathcal{H}, \mathbb{P}^x)$, the martingale $(\hat{M}_{T\wedge t})_{t\in\mathbb{R}_+}$ is L^2-bounded and, therefore, is uniformly integrable. By Theorem V.5.14, then, $\hat{\text{M}}$ is a Doob martingale on $[0, T]$, which implies that $\mathbb{E}^x \hat{M}_T = \mathbb{E}^x \hat{M}_0 = 0$ as needed. □

Infinitesimal generator

This is an extension of the operator G defined by 2.21. We keep the same notation, but we define it anew.

Let $\mathcal{D}_G$ be the collection of functions $f : E \mapsto \mathbb{R}$ for which the limit

2.28 $$Gf(x) = \lim_{t\downarrow 0} \frac{1}{t} \left[P_t f(x) - f(x)\right]$$

exists for every x in E. Then, G is called the *infinitesimal generator* of X, and $\mathcal{D}_G$ is called its *domain.*

2.29 Lemma. *Let $f \in \mathcal{C}_K^2$. Then, $f \in \mathcal{D}_G$, and the limit in 2.28 is given by 2.21.*

Proof. Let $f \in \mathcal{C}_K^2$, define Gf by 2.21. By Corollary 2.26, then, Gf is continuous and bounded, which implies that $P_s Gf(x) = \mathbb{E}^x\, Gf \circ X_s$ goes to $Gf(x)$ as $s \to 0$; this is by the bounded convergence theorem and the continuity of X. Thus, from the formula of 2.26, $Gf(x)$ is equal to the limit on the right side of 2.28. □

Forward and backward equations

2.30 Theorem. *Let $f \in \mathcal{C}_K^2$. Then, $f \in \mathcal{D}_G$, Gf is given by 2.21, and*

2.31 $$\frac{d}{dt} P_t f(x) = P_t G f(x), \qquad x \in E,\; t \in \mathbb{R}_+.$$

Moreover, for f in $\mathcal{C}_K^2$ again, $P_t f \in \mathcal{D}_G$ and, with G as in 2.28,

2.32 $$\frac{d}{dt} P_t f(x) = G P_t f(x), \qquad x \in E,\; t \in \mathbb{R}_+.$$

2.33 Remark. The equation 2.31 is called *Kolmogorov's forward* equation, because G is in front of P_t. By the same logic, 2.32 is called *Kolmogorov's backward equation.* Writing $u(t,x)$ for $P_t f(x)$ for fixed f, the backward equation can be re-written as

$$\frac{d}{dt}\, u = Gu, \qquad u(0,x) = f(x),$$

with the understanding that G applies to the spatial variable, that is, to $x \mapsto u(t,x)$. This sets up a correspondence between diffusions and partial differential equations, since functions in $\mathcal{D}_G$ can be approximated by sequences of functions in $\mathcal{C}^2_K$.

Proof. The first statement is mostly in Lemma 2.29 and Corollary 2.26: Let $f \in \mathcal{C}^2_K$. Then, Gf is given by 2.21, belongs to $\mathcal{D}_G$, and $s \mapsto Gf \circ X_s$ is continuous and bounded. Thus, by the bounded convergence theorem, $s \mapsto P_s G f(x) = \mathbb{E}^x\; Gf \circ X_s$ is continuous and bounded. Hence, in the equation for $P_t f$ given in Corollary 2.26, the integral on the right side defines a differentiable function in t; and, taking derivatives on both sides yields 2.31.

For f in $\mathcal{C}^2_K$, we have just shown that $t \mapsto P_t f(x)$ is differentiable. Thus, since $P_s P_t = P_{t+s}$, the limit

$$\begin{aligned} GP_t f(x) &= \lim_{s\to 0} \frac{1}{s}\, [P_s P_t f(x) - P_t f(x)] \\ &= \lim_{s\to 0} \frac{1}{s}\, [P_{t+s} f(x) - P_t f(x)] = \frac{d}{dt}\; P_t f(x) \end{aligned}$$

exists, that is, $P_t f \in \mathcal{D}_G$ and 2.32 holds. □

Potentials, resolvent

Let $f \in \mathcal{E}_b$ and $p > 0$. By the continuity of X, the mapping $(t,\omega) \mapsto X_t(\omega)$ is measurable relative to $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{H}$ and $\mathcal{B}(\mathbb{R}^d)$. Thus, the following defines a function in $\mathcal{E}_b$:

2.34 $$U_p f(x) = \mathbb{E}^x \int_0^\infty dt\; e^{-pt}\; f \circ X_t = \int_0^\infty dt\; e^{-pt}\; P_t f(x), \quad x \in E.$$

The function $U_p f$ is called the p-potential of f, and U_p is called the p-potential operator, and the family $(U_p)_{p>0}$ is called the *resolvent* of (P_t) or of X. Of course, 2.34 makes sense for f in $\mathcal{E}_+$ and $p \geq 0$ as well. The next theorem relates the resolvent to the infinitesimal generator: the operators U_p and $p - G$ are inverses of each other.

2.35 Theorem. *For $p > 0$ and $f \in \mathcal{C}^2_K$,*

$$U_p(p-G)f = (p-G)\, U_p f = f.$$

Proof. Fix p and f such. From Corollary 2.26, then,

$$pU_pf = \int_0^\infty dt\ pe^{-pt}\ f + \int_0^\infty dt\ pe^{-pt} \int_0^t ds\ P_s\ Gf = f + U_pGf$$

by a change in the order of integration over s and t. Thus, $U_p(pf - Gf) = f$. For the other claim, we start by noting that, since $P_sP_t = P_{s+t}$ and f is bounded,

$$P_sU_pf = \int_0^\infty dt\ e^{-pt}\ P_{s+t}\ f = e^{ps}\ U_pf - e^{ps} \int_0^s dt\ e^{-pt}\ P_tf.$$

In the rightmost integral, the integrand goes to f as $t \to 0$. Thus,

$$\begin{aligned} GU_pf &= \lim_{s\to 0} \frac{1}{s} (P_sU_pf - U_pf) \\ &= \lim_{s\to 0} \frac{e^{ps}-1}{s} U_pf - \lim_{s\to 0} \frac{1}{s}\ e^{ps} \int_0^s dt\ e^{-pt}\ P_tf = pU_pf - f. \end{aligned}$$

Thus, $(p - G)U_pf = f$ as well. □

Interpretations

Fix f in $\mathcal{E}_b$ and $p > 0$. Let T_p be a random variable having the exponential distribution with parameter p. Suppose that T_p is independent of X. Since $\mathbb{P}\{T_p > t\} = e^{-pt}$, we may express 2.34 as

2.36 $$U_pf(x) = \mathbb{E}^x \int_0^\infty dt\ 1_{\{T_p>t\}} f \circ X_t = \mathbb{E}^x \int_0^{T_p} dt\ f \circ X_t,$$

which is the expected earnings during $(0, T_p)$ if the rate of earnings is $f(y)$ per unit of time spent at y. A related interpretation is that

2.37 $$pU_pf(x) = \mathbb{E}^x \int_0^\infty dt\ pe^{-pt}\ f \circ X_t = \mathbb{E}^x\ f \circ X_{T_p},$$

and, equivalently, writing $P_{T_p}f(x)$ for $g \circ T_p$ with $g(t) = P_tf(x)$,

2.38 $$pU_pf(x) = \mathbb{E}\ P_{T_p}\ f(x).$$

These show, in particular, that pU_p is a Markov kernel on $(E, \mathcal{E})$. Moreover, noting that $T = pT_p$ is standard exponential, and since $\frac{1}{p}T \to 0$ as $p \to \infty$, it follows from 2.38 that

2.39 $$\lim_{p\to\infty} pU_p\ f(x) = f(x)$$

provided that $\lim_{t\to 0} P_tf(x) = f(x)$, for instance, if f is continuous in addition to being bounded.

Resolvent equation

2.40 THEOREM. *For $p > 0$ and $q > 0$, we have $U_pU_q = U_q\ U_p$, and*

2.41 $$U_p + p\ U_p\ U_q = U_q + q\ U_q\ U_p.$$

Proof. Let $f \in \mathcal{E}_b$. Let T_p and T_q be independent of each other and of X, both exponential variables, with respective parameters p and q. Since $P_sP_t = P_tP_s$, it follows from 2.38 that

$$pq\ U_pU_qf(x) = \mathbb{E}\ P_{T_p}\ P_{T_q}\ f(x) = \mathbb{E}\ P_{T_q}P_{T_p}\ f(x) = qp\ U_qU_pf(x),$$

that is, $U_p\ U_q = U_q\ U_p$. To show the resolvent equation 2.41, we start with the ordinary Markov property:

$$\begin{aligned} \mathbb{E}^x \int_0^{s+t} du\ f \circ X_u &= \mathbb{E}^x \int_0^s du\ f \circ X_u + \mathbb{E}^x \int_0^t du\ f \circ X_{s+u} \\ &= \mathbb{E}^x \int_0^s du\ f \circ X_u + \int_E P_s(x, dy)\ \mathbb{E}^y \int_0^t du\ f \circ X_u. \end{aligned}$$

Since T_p and T_q are independent of X, we may replace s with T_p, and t with T_q. Then, using the interpretations 2.36 and 2.38, we get

$$\mathbb{E}^x \int_0^{T_p+T_q} du\ f \circ X_u = U_pf(x) + p\ U_pU_qf(x).$$

This proves 2.41 since $T_p + T_q = T_q + T_p$. □

Killing the diffusion

This is to describe an operation that yields an absorbing Markov process that coincides with X over an initial interval of time. Here X is the diffusion (described by Theorem 2.5 and examined above) with state space $E = \mathbb{R}^d$.

Let k be a positive Borel function on E. Let T be independent of the process X and have the standard exponential distribution (with mean 1). Define, for t in $\mathbb{R}_+$ and ω in Ω,

2.42 $$\hat{X}_t(\omega) = \begin{cases} X_t(\omega) & \text{if } T(\omega) > \int_0^t ds\ k \circ X_s(\omega), \\ \partial & \text{otherwise,} \end{cases}$$

where ∂ is a point outside E. This defines a stochastic process $\hat{X}$ with state space $\bar{E} = E \cup \{\partial\}$. We think of ∂ as the cemetery; it is a trap, and

2.43 $$\zeta = \inf \left\{ t \in \mathbb{R}_+ : \hat{X}_t = \partial \right\}$$

is the time X is killed. It follows from 2.42 and the assumptions on T that, with $\exp_{-x} = e^{-x}$ and $\mathcal{G}^0_\infty = \sigma\{X_s = s \in \mathbb{R}_+\}$,

2.44 $$\mathbb{P}\left\{\zeta > t \,|\, \mathcal{G}^0_\infty\right\} = \exp_- \int_0^t ds\ k \circ X_s.$$

Thus, in the language of Chapter VI, the particle X is killed at the time ζ of first arrival in a conditionally Poisson process with random intensity process $k \circ X$. It is common to refer to $\hat{X}$ as the process obtained from X by killing X at the rate $k(x)$ when at x.

The process $\hat{X}$ is Markov with state space $(\bar{E}, \bar{\mathcal{E}})$; its living space is $(E, \mathcal{E})$. We adopt the conventions 2.18 regarding the trap ∂; recall that every f in $\mathcal{E}_+$ is extended onto $\bar{E}$ by setting $f(\partial) = 0$. Thus, the transition function of $\hat{X}$ is determined by

2.45
$$\begin{aligned}\hat{P}_t f\,(x) &= \mathbb{E}^x\; f \circ \hat{X}_t \\ &= \mathbb{E}^x\; f \circ X_t\; 1_{\{\zeta > t\}} = \mathbb{E}^x\,(f \circ X_t)\left(\exp_- \int_0^t ds\; k \circ X_s\right),\end{aligned}$$

with f in $\mathcal{E}_+$ and t in $\mathbb{R}_+$ and x in E. The Markov property of $\hat{X}$ implies that $(\hat{P}_t)$ is a transition semigroup. Each $\hat{P}_t$ is a sub-Markov kernel on $(E, \mathcal{E})$; the defect $1 - \hat{P}_t(x,\ E)$ being $\mathbb{P}^x\{\zeta \le t\}$. The following relates $(\hat{P}_t)$ to (P_t).

2.46 Proposition. *Let $t \in \mathbb{R}_+,\ x \in E,\ f \in \mathcal{E}_+$. Then,*

2.47
$$P_t f(x) = \hat{P}_t f(x) + \int_0^t ds \int_E \hat{P}_s\,(x, dy)\;\; k(y)\;\, P_{t-s} f(y).$$

Proof. We condition on whether killing occurs before or after time t, and we use the Markov property of X:

$$\begin{aligned}P_t f(x) &= \mathbb{E}^x\; f \circ X_t\; 1_{\{\zeta > t\}} + \mathbb{E}^x\; f \circ X_t\; 1_{\{\zeta \le t\}} \\ &= \mathbb{E}^x\; f \circ \hat{X}_t + \int_{[0,t]\times E} \mathbb{P}^x\,\{\zeta \in ds, X_\zeta \in dy\}\, P_{t-s}\; f(y).\end{aligned}$$

This yields the claim via 2.45 and the observation that

$$\mathbb{P}^x\,\{\zeta \in ds,\ X_\zeta \in dy\} = ds\;\hat{P}_s(x, dy)\; k(y).$$ □

Let $(\hat{U}_p)$ denote the resolvent of the semigroup $(\hat{P}_t)$, and recall the resolvent (U_p) of (P_t). Taking Laplace transforms on both sides of 2.47 we get

2.48
$$U_p f(x) = \hat{U}_p f(x) + \int_E \hat{U}_p(x, dy)\; k(y)\; U_p f(x).$$

We use this to obtain the generator $\hat{G}$ corresponding to $(\hat{P}_t)$ from the generator G of (P_t); see 2.21 and 2.28; in particular, $\hat{G}$ is defined by 2.28 from $(\hat{P}_t)$.

Let $f \in \mathcal{C}_K^2$. Recall Theorem 2.35 to the effect that $f = U_p(p - G)f$. Thus, in view of 2.48,

$$f = \hat{U}_p(p - G)\; f + \hat{U}_p(kf) = \hat{U}_p(p - \hat{G})f$$

where

2.49 $$\hat{G}f(x) = Gf(x) - k(x)f(x), \qquad f \in \mathcal{C}_K^2,\ x \in E.$$

In words, every f in $\mathcal{C}_K^2$ is in the domain of $\hat{G}$, and $\hat{G}$ is related to G and k through 2.49. Considering the relationship 2.28 for $\hat{G}$ and $(\hat{P}_t)$, and considering the formula 2.45 for $\hat{P}_t$, we obtain the following. This is known as *Feynman-Kac formula.*

2.50 PROPOSITION. *Let $f \in \mathcal{C}_K^2$ and put*

$$u(t,x) = \mathbb{E}^x\,(f \circ X_t)\left(\exp_- \int_0^t ds\ k \circ X_s\right), \qquad t \in \mathbb{R}_+, x \in E.$$

Then, u satisfies the partial differential equation

$$\frac{\partial}{\partial t}\,u = Gu - ku, \qquad u\,(0,\cdot) = f.$$

Proof of existence and uniqueness

We start the proof of Theorem 2.13 with a lemma on some approximations. We omit time subscripts that are variables of integration. Condition 2.4 is in force throughout.

2.51 LEMMA. *Let Y and Z be continuous processes with state space $E = \mathbb{R}^d$, put*

$$A_t = \int_0^t (u_0 \circ Y - u_0 \circ Z)\ ds,\ \ M_t = \sum_{n=1}^m \int_0^t (u_n \circ Y - u_n \circ Z)\ dW^n.$$

Then,

$$\mathbb{E}\sup_{s\le t} |A_s + M_s|^2 \le (2t + 8m)\,c^2 \int_0^t \mathbb{E}\,|Y - Z|^2\ ds.$$

Proof. By ordinary considerations, using the Lipschitz condition 2.4,

$$\mathbb{E}\,|A_t|^2 \le t\int_0^t |u_0 \circ Y - u_0 \circ Z|^2\ ds \le tc^2 \int_0^t \mathbb{E}\,|Y - Z|^2\ ds.$$

Applying the rule 7.6 to each component of the d-dimensional martingale M, recalling that the W^n are independent, we get

$$\mathbb{E}\,|M_t|^2 = \sum_{n=1}^m \int_0^t |u_n \circ Y - u_n \circ Z|^2\ ds \le mc^2 \int_0^t \mathbb{E}\,|Y - Z|^2\ ds,$$

where the last step used the Lipschitz condition. From these, we obtain

$$\mathbb{E}\sup_{s\le t} |A_s|^2 \le tc^2 \int_0^t \mathbb{E}\,|Y - Z|^2\ ds,$$

$$\mathbb{E}\sup_{s\le t} |M_s|^2 \le 4\ \mathbb{E}\,|M_t|^2 \le 4\ mc^2 \int_0^t \mathbb{E}\,|Y - Z|^2\ ds,$$

where we used the Doob-Kolmogorov inequality in the last line. These two last expressions yield the lemma. □

2.52 *Proof of Theorem 2.13*. Consider the equation 2.5 with $X_0 = x$ for some fixed x in E. We define a sequence of continuous processes $X^{(k)}$ by setting $X_t^{(0)} = x$ for all t, and by letting

2.53 $$X_t^{(k+1)} = x + \int_0^t u_0 \circ X_s^{(k)}\ ds + \sum_{n=1}^{m} \int_0^t u_n \circ X_s^{(k)}\ dW_s^n$$

for $k \geq 0$. Then, $X^{(k+1)} - X^{(k)} = A + M$, where A and M are as in the preceding lemma with $Y = X^{(k)}$ and $Z = X^{(k-1)}$. Fix τ in $\mathbb{R}_+$ and put $\alpha = (2\tau + 8\,m)c^2$. It follows from the lemma that

2.54 $$\mathbb{E} \sup_{s \leq t} \left| X_s^{(k+1)} - X_s^{(k)} \right|^2 \leq \alpha \int_0^t \mathbb{E} \left| X_s^{(k)} - X_s^{(k-1)} \right|^2\ ds$$

for every $t \leq \tau$ and $k \geq 1$. Whereas, by the lemma again, this time with $Y = X^{(1)}$ and $Z = X^{(0)} = x$,

2.55 $$\mathbb{E} \left| X_t^{(1)} - X_t^{(0)} \right|^2 = \mathbb{E} \left| u_0(x)t + \sum_{n=1}^{m} u_n(x) W_t^n \right|^2 = |u_0(x)|^2\ t^2 + \sum_{n=1}^{m} |u_n(x)|^2\, t \leq \beta$$

where $\beta = (\tau^2 + m\tau)\ c^2 (1 + |x|)^2$ in view of Condition 2.4.

We put the bound 2.55 into 2.54 with $k = 1$, put the resulting inequality back into 2.54 with $k = 2$, and continue recursively. We get

$$\mathbb{E} \sup_{s \leq \tau} \left| X_s^{(k+1)} - X_s^{(k)} \right|^2 \leq \beta\ \alpha^k \tau^k / k!,$$

which, via Markov's inequality, yields

2.56 $$\mathbb{P} \left\{ \sup_{s \leq \tau} \left| X_s^{(k+1)} - X_s^{(k)} \right|^2 > \frac{1}{2^k} \right\} \leq \beta\, (4\alpha\tau)^k / k!.$$

The right side is summable over k. By the Borel–Cantelli lemma, then, there is an almost sure event Ω_τ such that, for every ω in Ω_τ, the sequence $\left(X_t^{(k)}(\omega) \right)_{k \in \mathbb{N}}$ is convergent in $E = \mathbb{R}^d$ uniformly for t in $[0, \tau]$. We define $X_t(\omega)$ to be the limit for ω in Ω_τ and put $X_t(\omega) = x$ for all other ω.

It follows from the uniformity of convergence and the continuity of $X^{(k)}$ that X is continuous on $[0, \tau]$. It follows from 2.53 that X satisfies the equation 2.5 for $t \leq \tau$. And, τ is arbitrary.

There remains to show the uniqueness. Let X and $\hat{X}$ be solutions to 2.5 with $X_0 = \hat{X}_0 = x$. Then, $X - \hat{X} = A + M$ in the notation of Lemma 2.51 with $Y = X$ and $Z = \hat{X}$. Thus, for fixed τ in $\mathbb{R}_+$, we have

$$\mathbb{E} \sup_{s \le t} \left| X_s - \hat{X}_s \right|^2 \le (2\tau + 8m)\, c^2 \int_0^t \mathbb{E} \left| X_s - \hat{X}_s \right|^2 \; ds$$

for all $t \le \tau$. It now follows from Gronwall's inequality (see Exercise 2.70) that the left side vanishes. Thus, almost surely, $X_t = \hat{X}_t$ for all $t \le \tau$; and τ is arbitrary. □

Dependence on the initial position

Let $X_t(\omega, x)$ denote the position $X_t(\omega)$ when $X_0 = x$. The next proposition shows that the dependence of X_t on x is continuous in the L^2-space of $(\Omega, \mathcal{H}, \mathbb{P})$ and, hence, in probability.

2.57 PROPOSITION. *For each t in $\mathbb{R}_+$,*

$$\lim_{x \to y} \mathbb{E} \left| X_t(x) - X_t(y) \right|^2 = 0.$$

Proof. Fix x and y in E. Note that $X_t(x) - X_t(y) = x - y + A_t + M_t$ in the notation of Lemma 2.51 with $Y_t = X_t(x)$ and $Z_t = X_t(y)$. Thus with fixed $\tau < \infty$ and $\alpha = (2\tau + 8m)c^2$, we have

$$\mathbb{E} \left| X_t(x) - X_t(y) \right|^2 \le 2 \left| x - y \right|^2 + 2\alpha \int_0^t \mathbb{E} \left| X_s(x) - X_s(y) \right|^2 \; ds$$

for all $t \le \tau$. Via Gronwall's inequality (see 2.70), this implies that

$$\mathbb{E} \left| X_t(x) - X_t(y) \right|^2 \le 2 \left| x - y \right|^2 \; e^{2\alpha t}, \qquad 0 \le t \le \tau.$$

The claim is immediate since τ is arbitrary. □

The preceding proposition implies that $X_t(x) \to X_t(y)$ in probability as $x \to y$ in $E = \mathbb{R}^d$. Thus, for $f : E \mapsto \mathbb{R}$ bounded and continuous, as $x \to y$,

2.58 $$P_t f(x) = \mathbb{E}\; f \circ X_t(x) \to \mathbb{E}\; f \circ X_t(y) = P_t f(y)$$

as in Theorem III.1.6. In other words, if f is bounded continuous, then so is $P_t f$ for each t. This is called the *Feller property* for (P_t); it will show that Itô diffusions form a subset of Hunt processes to be introduced in Section 5.

Exercises and complements

2.59 *Differential operator.* Specify the operator G defined by 2.21 for

a) the geometric Brownian motion of Example 2.9,

b) Ornstein-Uhlenbeck process of 2.10,

c) Brownian motion on the unit circle, Example 2.11.

2.60 *Ornstein-Uhlenbeck model.* Let V be the Ornstein–Uhlenbeck velocity process for the physical Brownian motion on $\mathbb{R}$; it satisfies

$$dV_t = -cV_t\ dt + b\ dW_t,$$

where $c > 0$ and $b > 0$ are constants, and W is Wiener. Then, the particle position process X satisfies $dX_t = V_t\ dt$. Write the equation of motion for the $\mathbb{R}^2$-valued motion (V_t, X_t). What is the corresponding generator G on $C^2_K(\mathbb{R}^2 \mapsto \mathbb{R})$? show that V and X are Gaussian processes assuming that $V_0 = v$ and $X_0 = x$ are fixed.
Hint: Write the solution for V, and use integration by parts to express V as an ordinary integral of W.

2.61 *Graphs.* Let X be an Itô diffusion satisfying 2.5. Put $Y_t = (t, X_t)$. Write Itô's formula for $f \circ Y_t$ with f in $C^2_K(\mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R})$. Show that Y is an Itô diffusion that satisfies

$$dY_t = \sum_{n=0}^{m} v_n \circ Y_t\ d\ Z^n_t,$$

where $Z^o_t = t$ and $Z^n_t = W^n_t$ for $n \geq 1$ and the vector fields $v_0, \ldots, v_m$ on $\mathbb{R}^{d+1}$ chosen appropriately. Specify the v_n.

2.62 *Applications to Brownian motion.* Here and in Exercises 2.63–2.66 below, X is a standard Brownian motion in $E = \mathbb{R}^d$ as in Example 2.22. For Borel subsets D of E define τ_D to be the time of exit from D, that is,

$$\tau_D = \inf\{t \in \mathbb{R}_+ :\ X_t \notin D\}.$$

Recall that, when $d = 1$ and $D = (-r,\ r)$, we have $\mathbb{E}^0\tau_D = r^2$. Show that, in general, $\mathbb{E}^x\tau_D < \infty$ for x in D, for D bounded.
Hint: If D is bounded, it is contained in an open ball of some radius $r < \infty$ centered at x, and that ball is contained in the cylinder $C = (x_1 - r, x_1 + r) \times \mathbb{R}^{d-1}$ if $x = (x_1, \ldots, x_d)$. Then, $\tau_D \leq \tau_C$, and $\mathbb{E}^x\tau_C = r^2$.

2.63 *Continuation.* Let D be a ball of radius r centered at the origin. Show that, for x in D,

$$\mathbb{E}^x\ \tau_D = \frac{r^2 - |x|^2}{d}$$

Hint: Use Dynkin's formula, Theorem 2.27, with f in $\mathcal{C}^2_K$ chosen so that $f(x) = |x|^2$ for x in D.

2.64 *Hitting of spheres.* For $r \geq 0$, let T_r be the time that Brownian motion X hits the sphere of radius r centered at the origin of $\mathbb{R}^d$. For $0 < q < |x| < r$, consider the probability

$$\alpha = \mathbb{P}^x\{T_q < T_r\},$$

that is, the probability that X exits $D = \{x \in E :\ q < |x| < r\}$ by touching the inner sphere.

a) For $d = 1$, X is a Doob martingale on $[0, \tau_D]$; use this to show that $\alpha = (r - |x|)/(r - q)$.

b) Let $d = 2$. Show that $\alpha = (\log r - \log|x|)/(\log r - \log q)$.

Hint: Let $f \in \mathcal{C}_K^2$ such that $f(x) = \log|x|$ for x in D. Use Dynkin's formula for such f and stopping time τ_D.

c) Let $d = 3$. Show that

$$\alpha = \left(r^{2-d} - |x|^{2-d}\right) \Big/ \left(r^{2-d} - q^{2-d}\right).$$

Hint: use Dynkin's formula with f in $\mathcal{C}_K^2$ such that $f|x| = |x|^{2-d}$ for x in D.

2.65 *Recurrence properties.* For $d = 1$, the Brownian motion X will hit every point y repeatedly without end; see Chapter VIII.

a) Let $d = 2$. Let $r \to \infty$ in 2.64b to show that

$$\mathbb{P}^x \{T_q < \infty\} = 1, \qquad 0 < q < |x|,$$

however small the disk of radius q is. Show, however, that

$$\mathbb{P}^x \{T_0 < \infty\} = 0, \qquad |x| > 0.$$

b) Let $d = 3$ and $0 < q < |x|$. Show that

$$\mathbb{P}^x \{T_q < \infty\} = (q/|x|)^{d-2}.$$

In summary, standard Brownian motion is "point recurrent" for $d = 1$, fails to be point recurrent but is "disk recurrent" for $d = 2$, and is "transient" for $d \geq 3$.

2.66 *Bessel processes with index $d \geq 2$.* Let $X = X_0 + W$, a standard Brownian motion in $\mathbb{R}^d$. Define $R = |X|$. Started at $x \neq 0$, the process X never visits the point 0; see 2.65a. Thus, the true state space for R is $(0, \infty)$. Since $d \geq 2$, the function $f : x \mapsto |x|$ is twice differentiable everywhere except the origin.

a) Use Itô's formula on $R = f \circ X$ to show that

$$dR_t = \frac{d-1}{2\ R_t}\ dt + \sum_{i=1}^{d} \frac{1}{R_t} X_t^i\ dW_t^i = \frac{d-1}{2R_t}\ dt + d\hat{W}_t$$

with an obvious definition for $\hat{W}$.

b) Show that $\hat{W}$ is a continuous local martingale with $\hat{W}_0 = 0$. Show, using 7.24, that $\hat{W}$ is a Wiener process (one-dimensional).

2.67 *Bessel with index 2.* Let $d = 2$ in the preceding exercise, and let $R_0 = r > 0$ be fixed. Define

$$Y_t = \log\ R_t, \qquad t \in \mathbb{R}_+.$$

a) Show that Y is a continuous local martingale with $Y_0 = \log r$.

b) Let τ be the time of exit for R from the interval $(p,\ q)$, where $0 < p < r < q$. Show that $\tau < \infty$ almost surely and that Y is bounded on $[0, \tau]$. Show that, as in 2.64b,

$$\mathbb{P}\{R_\tau = p\} = \frac{\log\ q - \log\ r}{\log\ q - \log\ p}.$$

2.68 *Continuation.* a) Show that $C = \langle Y, Y \rangle$ is given by

$$C_t = \int_0^t ds\ e^{-2Y_s}.$$

Use Theorem 7.27 to conclude that

$$Y_t = \log r + \tilde{W}_{C_t}$$

for some Wiener process $\tilde{W}$.

b) Solve the last equation for Y by expressing C_t in terms of $\tilde{W}$.

c) Conclude that the Bessel process R is a time-changed geometric Brownian motion: $R = Z_C$, where

$$Z_s = re^{\tilde{W}_s}, \qquad S_u = \int_0^u ds\,(Z_s)^2, \qquad C_t = \inf\{u > t:\ S_u > t\}.$$

2.69 *Bessel with index $d \geq 3$.* Take $d = 3$ in 2.66 and fix $R_0 = r > 0$.

a) Show that $Y = R^{2-d}$ is a local martingale.

b) Use Theorem 7.27 to show that $Y = Y_0 + \tilde{W}_C$, where $\tilde{W}$ is a Wiener process and $C = \langle Y, Y \rangle$. Thus,

$$Y_t = Z_{C_t},$$

where $Z_u = r^{2-d} + \tilde{W}_u,\ u \geq 0$, a Brownian motion.

c) Show that C is the functional inverse of S, where

$$S_u = (d-2)^{-2} \int_0^u (Z_s)^{(2-2d)/(d-2)}\ ds.$$

Conclude that R is a deterministic function of a random time-changed Brownian motion:

$$R_t = \left(Z_{C_t}\right)^{-1/(d-2)}.$$

2.70 *Gronwall's inequality.* Let f and g be positive continuous functions on $\mathbb{R}_+$. Suppose that, for some c in $\mathbb{R}$,

$$f(t) \leq g(t) + c \int_0^t f(s)ds.$$

show that, then,

$$f(t) \le g(t) + c \int_0^t e^{c(t-s)}\ g(s)\ ds.$$

Hint: First show that

$$e^{-ct} \int_0^t f(s)\ ds \le \int_0^t e^{-cs}\ g(s)\ ds.$$

3 Jump-Diffusions

Jump-diffusions are processes that are Itô diffusions between the jumps. The jump times form a point process, and the diffusions and jumps interact. The treatment uses notions from Itô diffusions and Poisson random measures.

The motion of interest is a right-continuous, piecewise continuous process with state space $E = \mathbb{R}^d$ and the attendant σ-algebra $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$. Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space, and $\mathcal{F}$ is an augmented right-continuous filtration. Adapted to $\mathcal{F}$, and independent of each other, W is an m-dimensional Wiener process and M is a standard Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ (with mean $Leb \times Leb$). In addition, X_0 is an E-valued random variable in $\mathcal{F}_0$; it will often be treated as a deterministic parameter. As before, $(x, H) \mapsto \mathbb{P}^x(H)$ is a regular version of the conditional probabilities 2.3.

The motion

The deterministic data are some vector fields $u_0, \ldots, u_m$ on E and a Borel function $j : E \times \mathbb{R}_+ \mapsto E$. The vector fields are as in the preceding section; the function j rules the jump sizes. The motion X of interest satisfies the following stochastic integral equation:

$$X_t = X_0 + \int_0^t a \circ X_s\ ds + \int_0^t b \circ X_s\ dW_s + \int_{[0,t]\times\mathbb{R}_+} M\,(ds, dv)\, j(X_{s-}, v).$$

3.1

Here, $a = u_0$ and b is the $d \times m$ matrix whose columns are $u_1, \ldots, u_m$; see 2.5–2.8 for various equivalent ways of expressing the first two integrals. Unless stated otherwise, the next condition is in force throughout.

3.2 Condition. a) Lipschitz condition 2.4 holds. b) There is a constant c in $\mathbb{R}_+$ such that $j(x, v) = 0$ for $v > c$ for all x in E.

This condition is sufficient to ensure the existence and uniqueness of a piecewise continuous solution to 3.1. The condition on j makes the last integral in 3.1 to be effectively over $[0, t] \times [0, c]$, which means that the jump times of X form a subset of the arrival times in a Poisson process with rate c. Between two successive jumps, the motion is an Itô diffusion satisfying

3.3 $$d\bar{X}_t = a \circ \bar{X}_t\ dt + b \circ \bar{X}_t\ dW_t.$$

In particular, the initial segment (before the first jump) of X coincides with the Itô diffusion $\bar{X}$ satisfying 3.3 with the initial condition $\bar{X}_0 = X_0$.

Obviously, if $j = 0$, then $X = \bar{X}$; this was the subject of the preceding section. At the other extreme is the case $u_0 = \cdots = u_m = 0$, in which case X is piecewise constant, that is, each path $X(\omega)$ is a step function; the reader is invited to take this to be the case on a first reading; we shall treat this special case toward the end of this section. In the middle, there is the case where $u_1 = \cdots = u_m = 0$, in which case X is piecewise deterministic.

Construction of X

The next theorem describes X under the standing condition 3.2. The proof is constructive and is helpful for visualizing the paths.

3.4 Theorem. *The equation 3.1 has a pathwise unique solution X that is piecewise continuous, right-continuous, and locally bounded.*

Proof. For fixed x in E and s in $\mathbb{R}_+$, let $(t, \omega) \mapsto \bar{X}_{s,t}(\omega, x)$ be the process that is the solution $\bar{X}_t$ to 3.3 with $t \geq s$ and $\bar{X}_s = x$. Under the Lipschitz condition 2.4, Theorem 2.13 applies, and $t \mapsto \bar{X}_{s,t}(\omega, x)$ is pathwise unique and continuous for almost every ω. Under condition 3.2b, the last integral in 3.1 is over $[0, t] \times [0, c]$ effectively. Since M is standard Poisson, its atoms over $\mathbb{R}_+ \times [0, c]$ can be labeled $(S_n,\ V_n)$ so that, for almost every ω,

3.5 $$0 < S_1(\omega) < S_2(\omega) < \cdots, \ \lim S_n(\omega) = +\infty.$$

By eliminating from Ω a negligible event, we assume that these properties (on $\bar{X}$ and M) hold for every ω.

Fix ω, put $S_0(\omega) = 0$, and suppose that $X_t(\omega)$ is specified for all $t \leq s$, where $s = S_n(\omega)$ for some $n \geq 0$. We proceed to specify it for t in $(s,u]$, where we put $u = S_{n+1}(\omega)$. Since M_ω has no atoms in $(s, u) \times [0, c]$, the equation 3.1 implies that

$$X_t(\omega) = \bar{X}_{s,t}(\omega, X_s(\omega)), \qquad s \leq t < u.$$

Since $t \mapsto \bar{X}_{s,t}(\omega, x)$ is continuous and bounded on the interval $[s,u]$, we have

$$X_{u-}(\omega) = \lim_{t \nearrow u} X_t(\omega) = \bar{X}_{s,u}(\omega, X_s(\omega)),$$

which point is in E. Now, 3.1 implies that

$$X_u(\omega) = X_{u-}(\omega) + j\left(X_{u-}(\omega), V_{n+1}(\omega)\right).$$

This completes the specification of $X_t(\omega)$ for $t \leq S_{n+1}(\omega)$, and therefore for all t in $\mathbb{R}_+$, by recursion, in view of 3.5. The other claims of the theorem are immediate from this construction. □

In the proceding proof, the times S_n are the arrival times of the Poisson process $t \mapsto M([0, t] \times [0, c])$. The proof shows that the path $X(\omega)$ is continuous except possibly at times $S_n(\omega)$. Generally, not every $S_n(\omega)$ is a jump

time for $X(\omega)$. See Exercise 3.88 for an example where X has at most finitely many jumps, and, in fact, there is a strictly positive probability that it has no jumps.

Markov and strong Markov properties

The process X has both. The proofs follow the same lines as those for diffusions: uniqueness of solutions to 3.1, and the Markov and strong Markov properties for W and M.

3.8 THEOREM. *For each time t, the process $\hat{X} = (X_{t+u})_{u\in\mathbb{R}_+}$ is conditionally independent of $\mathcal{F}_t$ given X_t; given that $X_t = y$, the conditional low of $\hat{X}$ is the same as the law of X under $\mathbb{P}^y$.*

Proof. Analogous to the proof of Theorem 2.16, we have

$$\hat{X}_u = \hat{X}_0 + \int_0^u a\circ\hat{X}_s\, ds + \int_0^u b\circ\hat{X}_s\, d\hat{W}_s + \int_{[0,u]\times\mathbb{R}_+} \hat{M}\,(ds, dv)\, j\left(\hat{X}_{s-}, v\right),$$

where $\hat{W} = (W_{t+u} - W_t)_{u\in\mathbb{R}_+}$ and $\hat{M} = \{M(B_t) : B \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)\}$ with $B_t = \{(t+u, z) : (u, z) \in B\}$. Note that $\hat{W}$ and $\hat{M}$ are independent of $\mathcal{F}_t$ and of each other, $\hat{W}$ is Wiener just as W, and $\hat{M}$ is Poisson just as M. Finally, the uniqueness shown in Theorem 3.4 implies that $\hat{X}$ is obtained from $(\hat{X}_0, \hat{W}, \hat{M})$ by the same mechanism as X is obtained from (X_0, W, M). Hence, the claim. □

The strong Markov property is shown next; the wording repeats Theorem 2.19; the conventions 2.18 are in force regarding the end of time.

3.9 THEOREM. *The process X is strong Markov: For every $\mathcal{F}$-stopping time T, the variable X_T is $\mathcal{F}_T$-measurable, and $\hat{X} = (X_{T+u})_{u\in\mathbb{R}_+}$ is conditionally independent of $\mathcal{F}_T$ given X_T; moreover, for y in E, on the event $\{X_T = y\}$, the conditional law of $\hat{X}$ given X_T is the same as the law of X under $\mathbb{P}^y$.*

Proof. The measurability claimed for X_T is via Theorem V.1.14 and the right-continuity of X, and the adaptedness to $\mathcal{F}$. On $\{T = +\infty\}$, the claims regarding the conditional law given $\mathcal{F}_T$ are immediate from the conventions 2.18. On $\{T < \infty\}$, the claims are proved as in the preceding proof: replace t with T, recall that $\hat{W}$ is again Wiener by the strong Markov property for W (see Theorem VII.3.10), and use the next lemma.

3.10 LEMMA. *Let T be an $\mathcal{F}$-stopping time. For ω in $\{T < \infty\}$, define $\hat{M}(\omega, B) = M(\omega, B_{T(\omega)})$, where $B_t = (t, 0) + B,\ B \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)$. Then, on $\{T < \infty\}$, the conditional law of $\hat{M}$ given $\mathcal{F}_T$ is the law of the standard Poisson random measure M.*

Proof. Define $h(v) = v^{-2}$ for $v > 0$. Note that the lebesgue integral of $h \wedge 1$ over $\mathbb{R}_+$ is equal to 2, a finite number. It follows, since M is Poisson, adapted to $\mathcal{F}$, and with mean Leb $\times$ Leb on $\mathbb{R}_+ \times \mathbb{R}_+$, that

$$Z_t = \int_{[0,t]\times(0,\infty)} M\,(ds,\ dv)\ \ h(v)$$

defines an increasing pure-jump Lévy process adapted to $\mathcal{F}$; in fact, Z is stable with index 1/2. By Theorem VII.3.10, then, Z is strong Markov: on the event $\{T < \infty\}$, the conditional law of $\hat{Z} = (Z_{T+u} - Z_T)_{u\in\mathbb{R}_+}$ given $\mathcal{F}_T$ is the same as the law of Z.

Since h is a homeomorphism of $(0,\infty)$ onto $(0,\infty)$, the measure M_ω and the path $Z(\omega)$ determine each other for almost every ω. Similarly, on $\{T < \infty\}$, the measure $\hat{M}_\omega$ and the path $\hat{Z}(\omega)$ determine each other. Obviously, $\hat{M}$ bears the same relationship to $\hat{Z}$, as M does to Z. Thus, the claim follows from the strong Markov property of Z. □

Lévy kernel for jumps

This is a transition kernel from $(E, \mathcal{E})$ into $(E, \mathcal{E})$. It gives the rates and effects of jumps. It is defined by, for x in E and B in $\mathcal{E}$,

3.11 $$L\,(x, B) = \text{Leb}\,\{v \geq 0 :\ j(x, v) \neq 0,\ x + j(x, v) \in B\}\,.$$

Note that $L(x, \{x\}) = 0$. In a sense to be made precise by the next theorem, $L(x, B)$ is the rate of jumps from x into B per unit of time spent at x. If X were a Lévy process with a Lévy measure λ for its jumps, then $j(x, v)$ would be free of x, and $L(x, B)$ would be equal to $\lambda(B - x)$. Hence, the term "Lévy kernel" for L.

Heuristically, then, the "rate" of jumps when at x is

3.12 $$k(x) = L(x, E) = \text{Leb}\,\{v \geq 0 :\ j(x, v) \neq 0\}\,,\ x \in E.$$

Clearly, k is a positive Borel function on E. The general theory allows $k(x) = +\infty$ for some or for all x; Condition 3.2b implies that k is bounded by the constant c of 3.2b, and then, L is a bounded kernel. The next theorem does not assume 3.2b.

3.13 THEOREM. *Let f be a positive Borel function on $E \times E$. Let F be a positive left-continuous process adapted to $\mathcal{F}$. Then, for every x in E,*

3.14 $$\mathbb{E}^x \sum_{s\in\mathbb{R}_+} F_s\ f\circ(X_{s-}, X_s)\ \ 1_{\{X_{s-}\neq X_s\}} = \mathbb{E}^x \int_{\mathbb{R}_+} ds\ F_s \int_E L(X_s, dy)\ f\circ(X_s, y).$$

3.15 REMARKS. a) The proof below will show that this theorem holds without condition 3.2; all that is needed is that X be right-continuous and have left-limits in E, and that X satisfy the equation 3.1.

b) On the left side of 3.14, the sum is over the countably many times s of jumps. On the right side, we may replace X_s by X_{s-} since the integration over s will wash away the difference; the result with X_{s-} is closer to intuition.

c) Take F to be the indicator of $[0,\ t]$ for some fixed t, and let f be the indicator of a Borel rectangle $A \times B$. The sum on the left side of 3.14 becomes the number $N_t(A \times B)$ of jumps, during $[0,\ t]$, from somewhere in A to somewhere in B. Thus,

3.16 $$\mathbb{E}^x N_t(A \times B) = \mathbb{E}^x \int_0^t ds\, (1_A \circ X_{s-})\, L\,(X_{s-}, B)\,.$$

This is the precise meaning of the heuristic phrase that $L(y, B)$ is the rate of jumps from y into B. The phrase is short for the statement that $s \mapsto (1_A \circ X_{s-}) L(X_{s-}, B)$ is the random intensity for the point process $s \mapsto N_s(A \times B)$ in the sense employed in Chapter IV, Section 6.

c) In particular, for the total number $N_t(E \times E)$ of jumps X makes during $[0,\ t]$, we see from 3.16 that

3.17 $$\mathbb{E}^x\ N_t(E \times E) = \mathbb{E}^x \int_0^t ds\ k \circ X_{s-} = \mathbb{E}^x \int_0^t ds\ k \circ X_s.$$

Proof. Since X satisfies 3.1, the sum on the left side of 3.14 is equal to

$$\int_{\mathbb{R}_+ \times \mathbb{R}_+} M(ds, dv)\ F_s\ \hat{f}\,(X_{s-}, X_{s-} + j(X_{s-}, v))$$

where $\hat{f}(y,\ z) = f(y,\ z)$ for $y \neq z$, and $\hat{f}(y,\ z) = 0$ for $y = z$. Here, the integrand is a process $(\omega, s, v) \mapsto G(\omega, s, v)$ that satisfies the predictability conditions of Theorem VI.6.2 on Poisson integrals: F is left-continuous, $s \mapsto X_{s-}$ is left-continuous, both are adapted to $\mathcal{F}$, and $\hat{f}$ and j are Borel. It follows from that theorem that the left side of 3.14 is equal to

$$\begin{aligned}\mathbb{E}^x & \int_{\mathbb{R}_+ \times \mathbb{R}_+} ds\ dv\ F_s\ \hat{f}\,(X_{s-}, X_{s-} + j\,(X_{s-}, v)) \\ & = \mathbb{E}^x \int_{\mathbb{R}_+} ds\ F_s \int_E L\,(X_{s-}, dy)\, f(X_{s-}, y),\end{aligned}$$

where we used the definition 3.11 of L to evaluate the integral over v. The last expression is equal to the right side of 3.14 since X_{s-} differs from X_s for only countably many s. □

Under Condition 3.2b, the kernel L is bounded, as $k(x) = L(x,\ E) \leq c$. We repeat this, and add a consequence:

3.18 Corollary. *The kernel L is bounded. For every bounded Borel function f on $E \times E$,*

$$m_t = \sum_{s \le t} f \circ (X_{s-}, X_s)\, 1_{\{X_{s-} \ne X_s\}} - \int_0^t ds \int_E L(X_s, dy)\ f(X_s, y),\ t \ge 0,$$

is an $\mathcal{F}$-martingale with $m_0 = 0$.

Proof. Fix $0 \le t < u$, fix f bounded and positive. For an event H in $\mathcal{F}_t$, put $F_s = 1_H\ 1_{(t,u]}(s)$. Then, F is left-continuous and adapted, and it follows from the preceding theorem that

$$\mathbb{E}^x\ 1_H \sum_{t<s\le u} f(X_{s-}, X_s)\, 1_{\{X_{s-} \ne X_s\}} = \mathbb{E}^x 1_H \int_t^u ds \int_E L(X_s, dy)\, f(X_s, y).$$

Since f is bounded, and L is a bounded kernel, the right side is real-valued; passing it to the left, we see that

$$\mathbb{E}^x\ 1_H \cdot (m_u - m_t) = 0.$$

That is, (m_t) is a martingale when f is bounded positive Borel. For f bounded Borel, the same conclusion holds obviously. □

Generator

Recall the Itô diffusion $\bar{X}$, which is the solution to 3.3 with $\bar{X}_0 = X_0$. Let $\bar{G}$ be its generator, that is, $\bar{G}f(x)$ is given by the right side of 2.21 for f in $\mathcal{C}_K^2 = C_K^2(E \mapsto \mathbb{R})$. We introduce (condition 3.2 is in force)

3.19 $$Gf(x) = \bar{G}f(x) + \int_E L(x, dy)\,[f(y) - f(x)]\,, \quad f \in \mathcal{C}_K^2.$$

This integro-differential operator is the *generator* for X:

3.20 Theorem. *For every f in $\mathcal{C}_K^2$,*

$$M_t = f \circ X_t - f \circ X_0 - \int_0^t ds\ Gf \circ X_s, \qquad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-martingale.

Proof. a) Fix f in $\mathcal{C}_K^2$. Put $T_0 = 0$, and let $T_1,\ T_2, \ldots$ be the successive jump times of X, defined recursively via $T_{n+1} = \inf\{t > T_n : X_{t-} \ne X_t\}$ with $n \ge 0$. On the event $\{T_n \le t < T_{n+1}\}$ consider the telescoping sum

3.21 $$\begin{aligned} f \circ X_t - f \circ X_0 &= \sum_{i=1}^n \left(f \circ X_{T_i} - f \circ X_{T_{i-}}\right) \\ &\quad + \sum_{i=1}^n \left(f \circ X_{T_{i-}} - f \circ X_{T_{i-1}}\right) + f \circ X_t - f \circ X_{T_n} \\ &= A + B + C. \end{aligned}$$

b) The term A is a sum over the jump times during $[0,t]$. By Corollary 3.18,

$$3.22 \quad A = \sum_{s \le t} (f \circ X_s - f \circ X_{s-}) = m_t + \int_0^t ds \int_E L(X_s, dy)\,[f(y) - f(X_s)]\,.$$

c) We now prepare to evaluate $B + C$. Let S and T be stopping times, chosen so that, on the event $\{T < \infty\}$, we have $S < T$ and X continuous over the interval (S, T). Thus, on $\{T < \infty\}$, the process X coincides over (S, T) with some diffusion satisfying 3.3; and, since that diffusion has the generator $\bar{G}$,

$$3.23 \qquad f \circ X_{T-} - f \circ X_S = \bar{m}_T - \bar{m}_S + \int_S^T ds\ \bar{G} f \circ X_s,$$

by Theorem 2.23, with $\bar{m}$ as the martingale on the right side of 2.24.

d) Apply 3.23 repeatedly, with $S = T_{i-1}$ and $T = T_i$ for $i = 1, \ldots, n$, and with $S = T_n$ and $T = t$. We see that, on the event $\{T_n \le t < T_{n+1}\}$

$$3.24 \qquad B + C = \bar{m}_t + \int_0^t ds\ \bar{G} f \circ X_s.$$

Finally, put 3.22 and 3.24 into 3.21, recall the definition 3.19 of G, and put $M = m + \bar{m}$. The result is the claim, since the union of the events $\{T_n \le t < T_{n+1}\}$ is the event $\{\lim\ T_n = +\infty\}$, and the latter event is almost sure in view of Condition 3.2b. □

Transition function, forward equation

The transition function (P_t) for X is defined, as usual, by

$$3.25 \qquad P_t f(x) = \mathbb{E}^x f \circ X_t, \qquad x \in E,\ f \in \mathcal{E}_+.$$

It follows from the preceding theorem that, for f in $\mathcal{C}_K^2$,

$$3.26 \qquad \begin{aligned} P_t f(x) &= f(x) + \mathbb{E}^x \int_0^t ds\ G f \circ X_s \\ &= f(x) + \int_0^t ds\ P_s G f(x), \end{aligned}$$

where the interchange of expectation and integration is justified by noting that Gf is bounded: For f in $\mathcal{C}_K^2$, Corollary 2.25 shows that $\bar{G}f$ is bounded continuous, and L is a bounded kernel under the standing condition 3.2.

The equation 3.26 is the integrated form of Kolmogorov's forward equation; see Corollary 2.26 and Theorem 2.30 for the diffusion case. Indeed, a formal differentiation of 3.26 yields the counterpart of the differential equations 2.31. The differentiability here, however, requires some continuity for the jump function j (in addition to Lipschitz continuity for the velocity fields u_n).

3.27 THEOREM. *Suppose that $x \mapsto j(x,v)$ is continuous for every v in $\mathbb{R}_+$. Then, for $f \in \mathcal{C}_K^2$, Gf is bounded and continuous, and*

3.28
$$\frac{d}{dt}P_t f = P_t G f.$$

Proof. Fix f in $\mathcal{C}_K^2$. As mentioned in Corollary 2.25, then, $\bar{G}f \in \mathcal{C}_K$ and thus bounded continuous. On the other hand, by the definition of L, and by 3.2b,

3.29
$$\int_E L(x,dy)\,[f(y)-f(x)] = \int_0^c dv \quad [f(x+j(x,v))-f(x)].$$

Since f is bounded continuous, and $x \mapsto j(x,v)$ is continuous by assumption, the integral on the right side yields a bounded continuous function (via the bounded convergence theorem). Adding 3.29 to $\bar{G}f$, we see that Gf is bounded continuous.

Consequently, by the right-continuity of X,

$$\lim_{s\to 0} P_{t+s}Gf(x) = \lim_{s\to 0} \mathbb{E}^x\ Gf\circ X_{t+s} = \mathbb{E}^x\ Gf\circ X_t = P_t Gf(x).$$

Hence, with the help of 3.26, we get

$$\lim_{u\to 0}\frac{1}{u}\,[P_{t+u}f(x) - P_t f(x)] = \lim_{u\to 0}\frac{1}{u}\int_0^u ds\ P_{t+s}Gf(x) = P_t Gf(x). \qquad \square$$

The first jump time

We return to the master equation 3.1. Define R to be the time of first jump:

3.30
$$R = \inf\{t > 0 : X_{t-} \neq X_t\}.$$

We show next that R is the lifetime of the diffusion $\bar{X}$ killed at the rate $k(x)$ when at x; recall $\bar{X}$ of 3.3 with $\bar{X}_0 = X_0$, and recall the notation $\exp_- x$ for e^{-x}.

3.31 LEMMA. $\mathbb{P}\{R > t|\bar{X}\} = \exp_- \int_0^t\ ds\ k\circ \bar{X}_s,\ t\in\mathbb{R}_+.$

Proof. Pick an outcome ω. Note that $R(\omega) > t$ if and only if $X_s(\omega) = \bar{X}_s(\omega)$ for all $s \le t$, which is in turn equivalent to having

$$M(\omega, D_\omega) = 0 \quad \text{for} \quad D_\omega = \left\{(s,v)\in\mathbb{R}_+\times\mathbb{R}_+ : s\le t, j\left(\bar{X}_{s-}(\omega), v\right)\neq 0\right\}.$$

The diffusion $\bar{X}$ is determined by W, and M is independent of W. Thus, since M is Poisson with mean $\mu = \text{Leb}\times\text{Leb}$,

$$\mathbb{P}\left\{R > t|\bar{X}\right\} = \mathbb{P}\left\{M(D) = 0|\bar{X}\right\} = e^{-\mu(D)}.$$

Finally, it follows from the definition 3.12 of k that

$$\mu(D_\omega) = \int_0^t\ ds\ k\circ\bar{X}_s(\omega). \qquad \square$$

3.32 REMARK. The preceding lemma is without conditions on the jump function. As a result, all values in $[0, \infty]$ are possible for R. When Condition 3.2b is in force, k is bounded and, thus, $R > 0$ almost surely. But R can take $+\infty$ as a value, that is, it is possible that there are no jumps; see Exercise 3.88 for an example.

3.33 PROPOSITION. *Let $x \in E$, $f \in \mathcal{E}_+$, $t \in \mathbb{R}_+$. Then,*

$$\mathbb{E}^x f \circ X_t \; 1_{\{R>t\}} = \mathbb{E}^x \; f \circ \bar{X}_t \; \exp_- \int_0^t ds \; k \circ \bar{X}_s,$$

where $\bar{X}$ is the diffusion that is the solution to 3.3 with $\bar{X}_0 = X_0 = x$.

Proof. On the left side, we may replace X_t with $\bar{X}_t$, since they are the same on the event $\{R > t\}$. Now, conditioning on $\bar{X}$ and applying the last lemma yield the claim. □

The preceding proposition establishes a connection to the Feynman-Kac formula discussed earlier. Define

3.34 $$\hat{P}_t f(x) = \mathbb{E}^x \; f \circ \bar{X}_t \; \exp_- \int_0^t ds \; k \circ \bar{X}_s, \quad x \in E, \; f \in \mathcal{E}_+, \; t \in \mathbb{R}_+,$$

which is the right side of the formula in the preceding proposition. Then, $(\hat{P}_t)$ is the sub-Markov transition semigroup of the Markov process $\hat{X}$ obtained from the diffusion $\bar{X}$ by killing the latter at the rate $k(x)$ when at x. See 2.42–2.50 for these matters, the semigroup, computational results for it, and the associated resolvant and generator. At this point, we regard $(\hat{P}_t)$ as known.

Regeneration at R

Heuristically, R is the killing time of $\bar{X}$. On the event that the killing succeeds, it occurs at the location $\bar{X}_R = X_{R-}$, and a new diffusion is born at the point X_R. We think of R as the time of first *regeneration* for X.

3.35 THEOREM. *For every x in E,*

$$\mathbb{P}^x\{R \in ds, \; X_{R-} \in dy, \; X_R \in dz\} = ds \; \hat{P}_s\,(x, dy)\, L(y, dz), \quad s \in \mathbb{R}_+, \; y \in E, \; z \in E.$$

Proof. The claim is equivalent to the more precise claim that

3.36 $$\mathbb{E}^x \; g \circ R \; f \circ (X_{R-}, X_R) = \int_{\mathbb{R}_+} ds \; g(s) \int_E \hat{P}_s\,(x, dy) \int_E L(y, dz) \; f(y, z)$$

for f positive Borel on $E \times E$ and g positive continuous and with compact support in $\mathbb{R}_+$. Fix f and g such. Let $F_s = g(s)\; 1_{\{s \le R\}}$ in Theorem 3.13. On the left side of 3.14, the sum consists of a single term, namely, $g \circ R$ $f \circ (X_{R-}, X_R)$; this is because the only jump time s in $[0, R(\omega)]$ is at $s = R(\omega)$

if $R(\omega) < \infty$, and $g \circ R(\omega) = 0$ if $R(\omega) = +\infty$ since g has compact support. Hence, 3.14 becomes

$$\mathbb{E}^x\ g{\circ}R\ f{\circ}(X_{R-}, X_R) = \mathbb{E}^x \int_{\mathbb{R}+} ds\ g(s)\ 1_{\{s \le R\}} \int_E L(X_s, dz)\ f(X_s, z).$$

On the right side, we may replace $\{s \le R\}$ with $\{R > s\}$ without altering the integral. The result is the right side of 3.36 in view of Proposition 3.33 and the definition 3.34. □

3.37 Remark. With k as defined by 3.12, we let K be a Markov kernel on $(E, \mathcal{E})$ that satisfies

$$L(x, B) = k(x)\ K(x, B), \qquad x \in E,\ B \in \mathcal{E}.$$

If $k(x) > 0$, then this defines $K(x, \cdot)$ uniquely; when $k(x) = 0$, it matters little how $K(x, \cdot)$ is defined, we choose $K(x, \{x\}) = 1$ in that case, note that $K(x, \{x\}) = 0$ if $k(x) > 0$. Replacing $L(y, dz)$ with $k(y)\ K(y,\ dz)$ yields the following heuristic explanation of the preceding theorem: Suppose that the particle is started at x. It survives until t and moves as a diffusion to arrive at dy; this has probability $\hat{P}_t(x, dy)$. Then, it gets killed during dt; this has probability $k(y)\ dt$. Finally, it is reborn in dz; this has probability $K(y, dz)$.

The process at its jumps

This is to describe X at its successive jumps. We do it under the standing condition 3.2, although much of this requires less.

Put $T_0 = 0$ and let $T_1,\ T_2, \ldots$ be the successive jump times, that is,

3.38 $$T_{n+1} = \inf\{t > T_n : X_{t-} \ne X_t\}, \qquad n \in \mathbb{N}.$$

we have $T_1 = R$ as in 3.30. Condition 3.2b implies that $R > 0$ almost surely, which implies, through the strong Markov property at T_n, that $T_{n+1} > T_n$ almost surely on $\{T_n < \infty\}$. Moreover, as the construction in Theorem 3.4 makes clear,

3.39 $$\lim\ T_n = +\infty \text{ almost surely;}$$

in other words, for almost every ω, for every t in $\mathbb{R}_+$ there is n (depending on t and ω) such that $T_n(\omega) \le t < T_{n+1}(\omega)$; it is possible that $T_{n+1}(\omega) = \infty$. Finally, with the conventions in 2.18, we define

3.40 $$Y_n = X_{T_n}, \qquad n \in \mathbb{N}.$$

The strong Markov property at the stopping times T_n implies that $(Y, T) = (Y_n, T_n)_{n\in\mathbb{N}}$ is a Markov chain with state space $\bar{E} \times \bar{\mathbb{R}}_+$. It has a special structure: For A in $\mathcal{E}$ and B in $\mathcal{B}_{\mathbb{R}+}$,

3.41 $$\mathbb{P}\{Y_{n+1} \in A,\ T_{n+1} - T_n \in B \,|\, \mathcal{F}_{T_n}\} = Q(Y_n,\ A \times B),$$

that is, given (Y_n), the conditional law of (T_n) is that of an increasing process with independent increments. The process (Y, T) is called a *Markov renewal chain*, and the times T_n are said to be regeneration times for X.

The kernel Q is specified by Theorem 3.35: take $n = 0$ in 3.41 and recall that $Y_0 = X_0$ and $T_1 - T_0 = R$. So,

3.42 $$Q(x, A \times B) = \int_B ds \int_E \hat{P}_s(x,\, dy)\, L\,(y,\, A) = \int_B\, ds\; \hat{P}_s L\,(x, A)\,.$$

It specifies the finite dimensional distributions of (Y, T) via 3.41. In particular, we have the iterative formula

3.43 $$\begin{aligned} Q^n(x, A \times B) &= \mathbb{P}^x \left\{Y_n \in A, T_n \in B\right\} \\ &= \int_{E\times\mathbb{R}_+} Q\,(x, dy, ds)\, Q^{n-1}\,(y, A \times (B - s)) \end{aligned}$$

for $n \geq 1$, and obviously, $Q^1 = Q$ and $Q^0(x, A \times B) = I(x, A)\delta_0(B)$. Computationally, in terms of the Laplace transforms

3.44 $$Q^n_p\,(x, A) = \int_{\mathbb{R}_+} Q^n\,(x, A \times ds)\, e^{-ps} = \mathbb{E}^x\; e^{-pT_n} 1_A \circ Y_n,$$

we see from 3.43 and 3.42 that $Q^1_p = Q_p$, and $Q^\circ_p = I$, and

3.45 $$Q_p = \hat{U}_p L, \qquad Q^n_p = (Q_p)^n, \qquad n \in \mathbb{N},\ p \in \mathbb{R}_+,$$

where $(\hat{U}_p)$ is the resolvent of the semigroup $(\hat{P}_t)$.

Transition function

This is to give an explicit formula for the transition function (P_t) in terms of the known objects $\hat{P}_t$, L, and Q^n. The result is made possible by 3.39, that is, by the fact (guaranteed by 3.2b) that there can be at most finitely many jumps during $[0,\, t]$.

3.46 Theorem. *Let $x \in E,\ f \in \mathcal{E}_+,\ t \in \mathbb{R}_+$. Then,*

$$P_t f(x) = \sum_{n=0}^{\infty} \int_{E\times[0,t]} Q^n(x, dy, ds)\ \hat{P}_{t-s} f(y).$$

Proof. By the strong Markov property at T_n,

$$\begin{aligned} &\mathbb{E}^x\ f \circ X_t\ 1_{\{T_n \leq t < T_{n+1}\}} \\ &\quad= \int_{E\times[0,t]} \mathbb{P}^x \left\{Y_n \in dy, T_n \in ds\right\} \mathbb{E}^y f \circ X_{t-s}\ 1_{\{R>t-s\}} \\ &\quad= \int_{E\times[0,t]} Q^n(x, dy, ds)\ \hat{P}_{t-s} f(y), \end{aligned}$$

where we used the definition of Q^n in 3.43, and Proposition 3.33 for the meaning of $\hat{P}_t$. Summing both sides over n in $\mathbb{N}$ yields the claimed formula since $\lim T_n = \infty$; see 3.39 *et seq.* □

The formula in the preceding theorem gives the unique solution of the integro-differential equation 3.28 for (P_t). The uniqueness follows from 3.39. By avoiding generators, we have also avoided the continuity condition on j used in Theorem 3.27.

Nevertheless, it may be useful to characterize (P_t) as the unique solution to something resembling the backward equations.

3.47 THEOREM. *Let $x \in E,\ f \in \mathcal{E}_+,\ t \in \mathbb{R}_+$. Then,*

$$P_t f(x) = \hat{P}_t f(x) + \int_0^t ds\ \hat{P}_s L P_{t-s} f(x).$$

Proof. We use the so-called renewal argument at the time R of first jump:

$$\begin{aligned} P_t f(x) &= \mathbb{E}^x f \circ X_t \cdot 1_{\{R>t\}} + \mathbb{E}^x f \circ X_t\ 1_{\{R \le t\}} \\ &= \hat{P}_t f(x) + \int_{[0,t]\times E} \mathbb{P}^x \{R \in ds, X_R \in dz\}\, \mathbb{E}^z\ f \circ X_{t-s} \\ &= \hat{P}_t f(x) + \int_0^t ds \int_E \hat{P}_s(x, dy) \int_E L(y, dz) P_{t-s} f(z) \end{aligned}$$

where we used the strong Markov property at R followed by the distribution provided by Theorem 3.35. □

Resolvent

Let (U_p) be the resolvent of (P_t), defined by 2.34, but for the present X and (P_t). Taking Laplace transforms on both sides of 3.26, assuming the same conditions hold, we get

3.48 $$p\, U_p f = f + U_p\, G f, \qquad t \in \mathcal{C}_K^2.$$

with the generator as defined by 3.19. It is usual to write this in the form $U_p(p - G)f = f$, thus emphasizing that U_p is the inverse of $p - G$.

We can avoid differentials by using the probabilistic derivations for (P_t). With $Q_p = \hat{U}_p L$ as in 3.45, we see from Theorems 3.46 and 3.47 that

3.49 $$U_p f = \sum_{n=0}^{\infty} (Q_p)^n\, \hat{U}_p f, \qquad f \in \mathcal{E}_+,$$

3.50 $$U_p f = \hat{U}_p\ f + Q_p\ U_p f, \qquad f \in \mathcal{E}_+.$$

Indeed, 3.49 is the unique solution to the integral equation 3.50.

3.51 PROPOSITION. *Let $p > 0$ and $f \in \mathcal{E}_{b+}$. Then, $g = U_p f$ given by 3.49 is the unique bounded solution to*

$$g = \hat{U}_p f + Q_p g.$$

Proof. Replace g on the right side with $\hat{U}_p f + Q_p g$ repeatedly. With the notation of 3.45, we get

$$\begin{aligned} g &= \hat{U}_p f + Q_p\, g \\ &= \hat{U}_p f + Q_p \hat{U}_p f + Q_p^2\, g = \cdots = \left(I + Q_p + \cdots + Q_p^n\right) \hat{U}_p f + Q_p^{n+1}\, g. \end{aligned}$$

In the last member, the first term is increasing to $U_p f$ given by 3.49. Thus, there remains to show that

$$\lim_{n\to\infty} Q_p^n\, g = 0.$$

for every g bounded positive, say, bounded by b. But, then,

$$Q_p^n\, g(x) \le b\; Q_p^n(x, E) = b\; \mathbb{E}^x\; e^{-pT_n} \to 0$$

as $n \to \infty$, because $p > 0$ and $T_n \to \infty$ almost surely. □

Simple step processes

These are pure-jump processes obtained by setting the vector fields $u_0, \ldots, u_m$ equal to zero, and keeping the condition 3.2b on the jump function j. Then, the Wiener processes W^n have no rôle to play, and the diffusion $\bar{X}$ satisfying 3.3 becomes motionless: $\bar{X}_t = X_0$ for all t. Thus, the process X is a right-continuous step process with a bounded Lévy kernel. In the next subsection, we shall discuss dropping the boundedness condition on the Lévy kernel and thus weakening the condition 3.2b.

The Markov and strong Markov properties remain unchanged. Theorem 3.13 on the Lévy kernel L stays the same, as is Corollary 3.18. The generator G is simpler: since $\bar{X}$ is motionless, $\bar{G}$ disappears;

3.52 $$Gf(x) = \int_E L\,(x, dy)\,[f(y) - f(x)]\,, \qquad f \in \mathcal{E}_b.$$

Theorem 3.20 becomes stronger; the claim holds for every f in $\mathcal{E}_b$. Similarly, 3.26 holds for every f in $\mathcal{E}_b$. Theorem 3.27 is stronger:

3.53 $$\frac{d}{dt} P_t f = P_t G f, \qquad f \in \mathcal{E}_b,$$

without the continuity assumption on j. That assumption was used in the proof to show that $s \mapsto Gf \circ X_s$ is bounded and right-continuous; we have boundedness via 3.52 and the boundedness of f and L; and $s \mapsto Gf \circ X_s$

is right-continuous because $s \mapsto X_s$ is a right-continuous step function (and thus $s \mapsto g \circ X_s$ is a right-continuous step function for arbitrary g).

Since $\bar{X}$ is motionless, killing it becomes simpler. For the time R of first jump, since $X_t = \bar{X}_t = X_0$ on $\{R > t\}$, Lemma 3.31 and 3.33 and 3.34 become

$$\mathbb{P}^x \{R > t\} = e^{-k(x)t}, \qquad \hat{P}_t f(x) = \mathbb{E}^x f \circ X_t \; 1_{\{R>t\}} = e^{-k(x)t} \; f(x),$$

and Theorem 3.35 becomes elementary (we drop $X_{R-} = X_0$ and use the notation of Remark 3.37)

$$\mathbb{P}^x \{R \in ds, X_R \in dy\} = (ds \; k(x) \; e^{-k(x)s})K(x, dy).$$

3.54 *Heuristics.* Suppose that the initial state is x. The particle stays there an exponential amount of time with parameter $k(x)$, and, independent of that amount, jumps to a new point y with probability $K(x, dy)$; then, it has a sojourn at y of exponential duration with parameter $k(y)$, followed by a jump to a new point chosen in accord with the probability law $K(y, \cdot)$; and so on. It is possible that, somewhere along its path, the particle lands at a point z with $k(z) = 0$; that z is a trap, and the particle stays there forever after.

For the transition function (P_t) and the resolvent (U_p), it is possible to give explicit and easy to interpret formulas. Heuristically, instead of the jumps of X, the idea is to concentrate on the kicks by the atoms of the Poisson M. Some kicks cause jumps, some not. If the particle is at x when it is kicked by an atom (s, v), it jumps to $x + j(x, v)$ if $j(x, v) \neq 0$, and it stays put if $j(x, v) = 0$. Following this reasoning as in the proof of Theorem 3.4, we obtain (with c as a bound for k)

3.55 $$P_t(x, A) = \sum_{n=0}^{\infty} \frac{e^{-ct}(ct^n)}{n!} Q^n(x, A), \qquad t \in \mathbb{R}_+, \; x \in E, \; A \in \mathcal{E},$$

where Q^n is the nth power of the Markov kernel Q on $(E, \mathcal{E})$ given by

3.56 $$Q(x, A) = \left(1 - \frac{k(x)}{c}\right) I(x, A) + \frac{k(x)}{c} \; K(x, A).$$

Thus, X has the form of a Markov chain subordinated to a Poisson process; see 1.26. The corresponding resolvent is

3.57 $$U_p = \frac{1}{c+p} \sum_{n=0}^{\infty} \left(\frac{c}{c+p}\right)^n \; Q^n, \qquad p > 0.$$

Step processes and extensions

We continue with the vector fields u_n set to zero, and we weaken condition 3.2b: instead of assuming that the Lévy kernel is bounded, we shall assume

only that it is finite. The classical examples are the processes with discrete state spaces.

The process X of interest has state space $E = \mathbb{R}^d$ as before. It is adapted to the filtration $\mathcal{F}$, it is right-continuous and has left-limits in E, and it satisfies

3.58 $$X_t = X_0 + \int_{[0,t]\times\mathbb{R}_+} M\,(ds, dv)\;\; j(X_{s-}, v),$$

with the same Poisson random measure M as before. Theorem 3.13 remains true (as remarked in 3.15a) with the Lévy kernel L defined by 3.11. We assume throughout that the following holds.

3.59 Condition. The Lévy kernel L is finite.

In other words, $k(x) = L(x, E) < \infty$ for every x in E. Further, we may assume that $j(x, v) \neq 0$ for $0 \leq v \leq k(x)$ only. Then, 3.58 is easier to visualize; see Exercise 3.84 also.

The process X has well-defined times $T_1, T_2, \ldots$ of the first jump, the second jump, ... It is a step process if and only if

3.60 $$T_\alpha = \lim_n \; T_n$$

is almost surely equal to $+\infty$. Otherwise, $t \mapsto X_t$ is a step function only over the interval $[0, T_\alpha)$. In either case, the evolution of X over $[0, T_\alpha)$ is as described in 3.54. The following two examples are instructive; see Exercises 3.90 and 3.91 as well.

3.61 Example. *Upward staircase.* Take $E = \mathbb{R}$. Let $D = \{x_0, x_1, \ldots\}$ where $0 = x_0 < x_1 < \cdots$ and $\lim x_n = 1$. Let $k(x) \in (0, \infty)$ for each x in D, and put $k(1) = 0$. If $x = x_n$ for some n and $v \leq k(x_n)$, then put $j(x, v) = x_{n+1} - x_n$; put $j(x, v) = 0$ for all other x and v.

If $X_0 = x_0 = 0$, then X stays at x_0 until T_1 and jumps to x_1, stay at x_1 until T_2 and jumps to x_2, and so on. The sojourn lengths $T_1,\ T_2 - T_1, \ldots$ are independent exponential variables with respective parameters $k(x_0),\ k(x_1), \ldots$. Their sum is the variable T_α defined by 3.60. So,

3.62 $$\mathbb{E}^0\; T_\alpha = \sum_{x\in D} \frac{1}{k(x)}.$$

If $\mathbb{E}^0 T_\alpha < \infty$, then $T_\alpha < \infty$ almost surely, and we let $X_t = 1$ for $t \geq T_\alpha$. We show next that, if $\mathbb{E}^0 T_\alpha = +\infty$, then $T_\alpha = +\infty$ almost surely and $X_t \in D$ for all t. In either case, we call X a staircase over $[0, T_\alpha)$ with steps at $x_0, x_1, \ldots$.

For the main computation, we use 3.16. For fixed n, let A consist of x_n, and B of x_{n+1}; then $N_t(A \times B)$ becomes the indicator of $\{X_t > x_n\}$, and 3.16 yields

3.63 $$\mathbb{P}^0\,\{X_t > x_n\} = \int_0^t \; ds\; \mathbb{P}^0\,\{X_s = x_n\}\;\; k(x_n).$$

Pass the factor $k(x_n)$ to the left side and note that $\{T_\alpha \le t\} \subset \{X_t > x_n\}$. Thus,

$$\frac{1}{k(x_n)} \mathbb{P}^0 \{T_\alpha \le t\} \le \int_0^t ds\ \mathbb{P}^0 \{X_s = x_n\}$$

Sum both sides over all n, note 3.62, and note that the sum of the right side is at most t. We get

$$\mathbb{E}^0 T_\alpha\ \mathbb{P}^0 \{T_\alpha \le t\} \le t.$$

We conclude that, if $\mathbb{E}^0 T_\alpha = +\infty$ then $\mathbb{P}^0\{T_\alpha \le t\} = 0$ for all t, which means that $T_\alpha = +\infty$ almost surely.

We re-state the essential content of the preceding example:

3.64 Lemma. *Let S be the sum of a countable independency of exponentially distributed random variables. If $\mathbb{E}S < \infty$ then $S < \infty$ almost surely; if $\mathbb{E}S = +\infty$ then $S = +\infty$ almost surely.*

3.65 Example. Let E, D, k, j be as in the last example, but with

3.66
$$\sum_{x \in D} \frac{1}{k(x)} = 1.$$

Let μ be a probability measure on D. We now describe a process that is a concatenation of staircases.

Started at $x = x_i$ for some i, the process is a staircase over $[0, T_\alpha)$ with steps at $x_i,\ x_{i+1}, \ldots$; in view of 3.66, we have $\mathbb{E}^x T_\alpha \le 1$, and thus $T_\alpha < \infty$ almost surely. At T_α, we deviate from Example 3.61: we choose the random variable X_{T_α} independent of $\mathcal{F}_{T_\alpha}$ according to the distribution μ on D. If X_{T_α} turns out to be x_j, we let X form a staircase over $[T_\alpha, T_{2\alpha})$ with steps at $x_j, x_{j+1}, \ldots$; note that $\mathbb{E}^x T_{2\alpha} \le 2$ and thus $T_{2\alpha} < \infty$ almost surely. We choose $X_{T_{2\alpha}}$ independent of $\mathcal{F}_{T_{2\alpha}}$ and with distribution μ again, and proceed to form another staircase over $[T_{2\alpha}, T_{3\alpha})$. And we repeat this over and over. The result is a process whose jump times can be ordered as

3.67 $T_1, T_2, \ldots, T_\alpha; T_{\alpha+1}, T_{\alpha+2}, \ldots, T_{2\alpha}; T_{2\alpha+1}, T_{2\alpha+2}, \ldots, T_{3\alpha}; \ldots.$

Each $T_{n\alpha}$ is the limit of a strictly increasing sequence of jump times; at each $T_{n\alpha}$ the process jumps from its left-limit 1 to its right-hand value $X_{T_{n\alpha}}$, the latter being independent of $\mathcal{F}_{T_{n\alpha}}$ and having the distribution μ. It follows that $T_{2\alpha} - T_\alpha, T_{3\alpha} - T_{2\alpha}, \ldots$ are independent and identically distributed, and, hence, $\lim_n T_{n\alpha} = +\infty$ almost surely. So, X_t is well-defined for every t in $\mathbb{R}_+$; the process X is right-continuous, is left-limited (as a process with state space E), and satisfies 3.58. Incidentally, this example shows that 3.58 can have many solutions.

We resume the treatment of the process X of 3.58–3.59, concentrating on the transition semigroups (P_t) and (P_t^*), where

3.68 $$P_t f(x) = \mathbb{E}^x\ f \circ X_t, \qquad P_t^* f(x) = \mathbb{E}^x\ f \circ X_t\ 1_{\{T_\alpha > t\}}.$$

Recall that for the time $R = T_1$ of first jump, we have

3.69 $Q\,(x, dy, ds) = \mathbb{P}^x\,\{X_R \in dy, R \in ds\} = ds\; k(x)e^{-k(x)s}\; K(x, dy),$

as with simple step processes; and $k(x) < \infty$ by assumption. Also as before, T_n is the time of nth jump and $Y_n = X_{T_n}$ for $n \in \mathbb{N}$, and

3.70 $Q^n(x, dy, ds) = \mathbb{P}^x\,\{Y_n \in dy, T_n \in ds\}\,, \qquad x, y \in E, s \in \mathbb{R}_+.$

Obviously, $Q^\circ(x, dy, ds) = I(x, dy)\,\delta_0(ds)$ and $Q^1 = Q$, and Q^n can be computed recursively via 3.43.

3.71 Proposition. *Let $x \in E,\ f \in \mathcal{E}_+,\ t \in \mathbb{R}_+$. Then,*

$$P_t^* f(x) = \sum_{n=0}^{\infty} \int_{E\times[0,t]} Q^n(x, dy, ds)\; e^{-k(y)(t-s)} f(y).$$

Proof. This is essentially as in the proof of Theorem 3.46: On the set $\{T_\alpha > t\}$ we have $T_n \le t < T_{n+1}$ for some n in $\mathbb{N}$. Hence,

$$\begin{aligned} P_t^* f(x) &= \sum_{n=0}^{\infty} \mathbb{E}^x\; f \circ X_t\; 1_{\{T_n \le t < T_{n+1}\}} \\ &= \sum_{n=0}^{\infty} \int_{E\times[0,t]} \mathbb{P}^x\,\{Y_n \in dy,\ T_n \in ds\}\; f(y)\; \mathbb{P}^y\,\{R > t - s\}\,, \end{aligned}$$

which is the claim. □

If $T_\alpha = +\infty$ almost surely, then $P_t^* = P_t$ for all t; see 3.68. A simple criterion for ensuring this condition is obtained from Lemma 3.64: Since the sojourn lengths $T_1, T_2 - T_1, \ldots$ are conditionally independent given (Y_n) and are conditionally exponential with parameters $k \circ Y_0,\ k \circ Y_1, \ldots$, Lemma 3.64 applies to the conditional law of (T_n) given (Y_n). The result is put next.

3.72 Proposition. *If $\sum_n 1/k \circ Y_n = +\infty$ almost surely, then $T_\alpha = +\infty$ almost surely and $P_t = P_t^*$ for all t.*

The preceding proposition is effective in a number of special situations: If k is bounded, then $T_\alpha = +\infty$ almost surely. If there is a recurrent point x for the chain Y, that is, if

$$\mathbb{P}^x\,\{Y_n = x \text{ for infinitely many } n\} = 1,$$

then $k \circ Y_n = k(x)$ for infinitely many n, and hence $T_\alpha = +\infty$ almost surely under $\mathbb{P}^x$. Similarly, if there is a recurrent set A (which Y visits infinitely often) and if k is bounded on A, then $T_\alpha = +\infty$ almost surely and $P_t^* = P_t$.

The next proposition is a summary of easy observations and a criterion for deciding whether $T_\alpha = +\infty$ almost surely.

3.73 PROPOSITION. *The following are equivalent:*

a) X is a step process, that is, $T_\alpha = +\infty$ almost surely.

b) $P_t^* = P_t$ for all t.

c) There exists $\tau > 0$ such that $P_\tau^*(x, E) = 1$ for all x in E.

d) For some (and therefore all) $p > 0$, the only solution to

3.74 $$h = Q_p h, \qquad 0 \le h \le 1, \qquad h \in \mathcal{E},$$

is $h = 0$; here $Q_p(x, A) = \frac{k(x)}{k(x)+p} K(x, A)$; see 3.69.

Proof. Obviously, $(a) \Leftrightarrow (b) \Rightarrow (c)$. To see that $(c) \Rightarrow (a)$, fix $\tau > 0$ such, that is, $P_\tau^* 1 = 1$. Then, $P_{s+\tau}^* 1 = P_s^* P_\tau^* 1 = P_s^* 1$ for all s. Replacing s with $\tau, 2\tau, \ldots$ we see that $P_{n\tau}^* 1 = 1$ for every n, which means that $\mathbb{P}^x\{T_\alpha > n\tau\} = 1$ for all x and n. Hence, $T_\alpha = +\infty$ almost surely.

Finally, we show that $(d) \Leftrightarrow a$. It follows from 3.69, 3.70, 3.43 that, for fixed $p > 0$,

$$h^*(x) = \mathbb{E}^x \; e^{-pT_\alpha} = \lim_n \mathbb{E}^x e^{-pT_n} = \lim_n \; Q_p^n 1(x),$$

where $Q_p^n = (Q_p)^n$. Thus, $Q_p \; h^* = \lim \; Q_p^{n+1} 1 = h^*$, that is, h^* is a solution to 3.74. Moreover, it is the maximal solution to it: if h is a solution, then $h \le h^*$, since

$$h = Q_p^n h \le Q_p^n 1 \to h^*.$$

Hence, $h^* = 0$ if and only if $h = 0$ is the only solution to 3.74. This shows that $(d) \Leftrightarrow (a)$, since $h^* = 0$ if and only if $T_\alpha = +\infty$ almost surely. □

The next theorem lists the backward equations for the derivatives of (P_t) and (P_t^*). We re-introduce the generator G:

3.75 $$Gf(x) = \int_E L\,(x, dy)\,[f(y) - f(x)], \qquad x \in E, f \in \mathcal{E}_b.$$

3.76 THEOREM. *Let $f \in \mathcal{E}_b$. We have the backward equations*

$$\frac{d}{dt} \; P_t f = G P_t f, \qquad \frac{d}{dt} P_t^* f = G P_t^* f.$$

These equations have a unique solution (and $P_t f = P_t^* f$ for all t) if and only if X is a step process.

The proof will be given together with the proof of the following, more comprehensive, result on the backward equations in integral form.

3.77 THEOREM. *Let $f \in \mathcal{E}_{b+}$. For bounded Borel $g : E \times \mathbb{R}_+ \mapsto \mathbb{R}_+$, consider the equation*

$$3.78 \qquad g(x,t) = e^{-k(x)t} f(x) + \int_0^t ds\, k(x)\, e^{-k(x)s} \int_E K(x,dy)\, g(y, t-s)$$

for x in E and t in $\mathbb{R}_+$. This equation holds for both g^o and g^*, where

$$g^o(x,t) = P_t f(x), \qquad g^*(x,t) = P_t^* f(x).$$

If X is a step process, then 3.78 has exactly one solution: $g = g^o = g^*$. Otherwise, the uniqueness fails, and g^* is the minimal solution.

3.79 REMARK. If X is not a step process, the backward equation characterizes $P_t^* f$ as the minimal solution, but does not specify $P_t f$. For instance, for Example 3.65, there are as many solutions as there are choices of μ. See Exercise 3.91 for the computation of $P_t f$.

Proof of 3.76, assuming 3.77. We re-write 3.78:

$$g(x,t) = e^{-k(x)t}\Big[f(x) + \int_0^t ds\; e^{k(x)s} \int_E L(x,dy)\, g(y,s)\Big].$$

On the right side, since g is bounded Borel, the integration over E yields a bounded Borel function, and the integration over $[0,\ t]$ yields a continuous function in t. Thus, on the left, $t \mapsto g(x,t)$ must be bounded continuous. We put this back into the right side: since $s \mapsto g(y,s)$ is bounded continuous, the integration over E yields a bounded continuous function in s, and the integration over $[0,\ t]$ yields a differentiable function in t. So, $t \mapsto g(x,t)$ is differentiable. Taking derivatives, we get

$$\frac{\partial}{\partial t} g(x,t) = -k(x) g(x,t) + \int_E L(x,dy)\; g(y,t) = \int_E L(x,dy)\,[g(y,t) - g(x,t)]\,.$$

Assuming Theorem 3.77, then, the functions g^o and g^* must satisfy the preceding; hence the backward equations of Theorem 3.76. The other assertion, on uniqueness, is immediate from Theorem 3.77. □

Proof of Theorem 3.77

We start by showing that g^* satisfies 3.78; showing the same for g^o is similar and simpler. We use the strong Markov property at the time R of first jump. Since $R < T_\alpha$ and $X_t = X_0$ on $\{R > t\}$,

$$\begin{aligned} P_t^* f(x) &= \mathbb{E}^x\ f \circ X_t\ 1_{\{R>t\}} + \mathbb{E}^x\ f \circ X_t\ 1_{\{R \le t\}} 1_{\{T_\alpha > t\}} \\ &= e^{-k(x)t}\ f(x) + \int_{E \times [0,t]} Q(x,dy,ds)\, P_{t-s}^* f(y), \end{aligned}$$

where Q is as in 3.69. This is the same as 3.78 for g^*.

Consider the solutions g to 3.78. We employ Laplace transforms with Q_p as in 3.74 (see also 3.69, 3.70, 3.43) and

$$g_p(x) = \int_0^\infty dt\ e^{-pt}\ g(x,t), \qquad f_p(x) = \frac{1}{k(x)+p} f(x).$$

Then, 3.78 becomes

$$g_p = f_p + Q_p\ g_p.$$

Replace g_p on the right side with $f_p + Q_p g_p$, and repeat n times. We get

$$g_p = \left(I + Q_p + \cdots + Q_p^n\right) f_p + Q_p^{n+1} g_p.$$

Since f is positive, the first term on the right is increasing in n; the limit is, in view of 3.71,

$$\sum_{n=0}^\infty Q_p^n f_p = \int_0^\infty dt\ e^{-pt} P_t^*\ f$$

Hence, g^* is the minimal solution to 3.78. The uniqueness has to do with

$$h_p = \lim_n\ Q_p^n\ g_p.$$

We note that h_p is bounded, positive, and satisfies

$$h_p = Q_p\ h_p.$$

Thus, the remaining assertions of the theorem follow from Proposition 3.73. □

Forward equations are more sensitive to whether X is a steps process. The next theorem shows that (P_t^*) satisfies the forward equation, but (P_t) does not. We list a needed result first; see 3.70 and 3.71.

3.80 Lemma. $\sum_{n=1}^\infty\ Q^n(x, dy, ds) = ds\ P_s^* L(x, dy),\ y \in E,\ s \in \mathbb{R}_+.$

Proof. In view of 3.70, what we need to show can be stated more precisely as

$$\mathbb{E}^x \sum_{n=1}^\infty\ 1_A \circ Y_n\ 1_{\{T_n \le t\}} = \int_0^t ds \int_E P_s^*(x, dy) L(y, A).$$

The left side is the same as the left side of 3.14 with $F_s = 1_{\{s \le t \wedge T_\alpha\}}$ and $f(x,y) = 1_A(y)$. Thus, the left side is equal to

$$\begin{aligned}&\mathbb{E}^x \int_0^\infty ds\ 1_{\{s \le t\wedge T_\alpha\}} \int_E L(X_s, dy)\ 1_A(y)\\ &= \mathbb{E}^x \int_0^t ds \quad 1_{\{T_\alpha > s\}} L(X_s, A) = \int_0^t ds \int_E P_s^*(x, dy) L(y, A),\end{aligned}$$

where we used the definition of P_t^* in 3.68 for the last equality. □

3.81 THEOREM. *Let $f \in \mathcal{E}_{b_+}$ and let the generator G be as in 3.75. Then, $P_t^* f$ satisfies the equation*

$$\frac{d}{dt}\ P_t^* f = P_t^* Gf$$

and is the minimal solution of it with $P_0^* f = f$. For (P_t) we have

$$\frac{d}{dt}\ P_t f \geq P_t Gf;$$

the equality holds if and only if X is a step process, and, then, $P_t f = P_t^* f$.

Proof. Combining Proposition 3.71 and Lemma 3.80, we have

3.82 $$P_t^* f(x) = e^{-k(x)t} f(x) + \int_0^t ds \int_E P_s^* L(x, dy) e^{-k(y)(t-s)} f(y).$$

By Theorem 3.76, this is differentiable. Taking derivatives on both sides we obtain

$$\begin{aligned}\tfrac{d}{dt}\ P_t^* f(x) &= -k(x)e^{-k(x)t} f(x) + P_t^* Lf(x) \\ &\quad - \textstyle\int_0^t ds\ P_s^* L(x, dy) e^{-k(y)(t-s)} k(y) f(y) \\ &= -k(x)e^{-k(x)t} f(x) + P_t^* Lf(x) - \left[P_t^*(kf)(x) - e^{-k(x)t} k(x) f(x)\right]\end{aligned}$$

where we used 3.82 at the last step. Hence, we have

3.83 $$\frac{d}{dt} P_t^* f = P_t^* Lf - P_t^*(kf) = P_t^* Gf.$$

For $P_t f$, we have differentiability by Theorem 3.76. And,

$$P_{t+s}\ f - P_t f = P_t(P_s f - f) \geq P_t\,(P_s^* f - f)$$

by 3.68 and positivity of f. Thus, using the boundedness of f,

$$\frac{d}{dt} P_t f = \lim_{s\to 0} \frac{P_{t+s} f - P_t f}{s} \geq \lim_{s\to 0} P_t \frac{P_s^* f - f}{s} = P_t\ Gf,$$

where we used 3.83 at the last step. The other assertions are repetitions of some claims in Theorem 3.77. □

Much of the foregoing are classical results for Markov processes with discrete state spaces. We have chosen the state space to be $E = \mathbb{R}^d$. Obviously, every countable set D with the discrete topology can be embedded in E, but our requirement of right-continuity for X leaves out an interesting class of processes which have discrete state spaces but permit discontinuities of the second kind; see the notes for this chapter.

Exercises and complements

3.84 *Lévy kernel and the jump function.* Let L be a finite kernel, that is, $k(x) = L(x,E) < \infty$ for every x in E, and take $E = \mathbb{R}$. Suppose that $L(x,\{x\}) = 0$ for each x and define $j(x,v)$ by

$$x + j(x,v) = \inf\{y \in \mathbb{R} \ : \ L(x,(-\infty,y]) > v\}$$

for v in $(0,k(x))$, and set $j(x,v) = 0$ for other v in $\mathbb{R}_+$.

a) Show that $y \mapsto L(x,(-\infty,y])$ and $v \mapsto x + j(x,v)$ are functional inverses of each other.

b) Show that L is the Lévy kernel defined from j by 3.11.

3.85 *Exponential decay with jumps.* Let $E = \mathbb{R}$, let $X_0 \in (0,\infty)$, and let X satisfy 3.1 with

$$a(x) = -cx, \qquad b(x) = 0, \qquad j(x,v) = xv \ 1_{(0,1)}(xv),$$

for $x > 0$. Note that X remains in $(0,\infty)$ forever. Plot the atoms of $M(\omega,\cdot)$ for a typical ω. Draw the path $t \mapsto X_t(\omega)$ corresponding to $M(\omega,\ \cdot)$ and with $X_0(\omega) = 1$.

3.86 *Continuation.* In the preceding exercise, replace j with

$$j(x,v) = j_0(v) 1_{(0,3)}(v)$$

for some increasing right-continuous function j_0 on (0,3). Describe the evolution of X with special attention to jump times, jump amounts, dependence and independence. Show that X satisfies the integral equation

$$X_t(\omega) = X_0(\omega) - c\int_0^t ds\ X_s(\omega) + Z_t(\omega), \qquad \omega \in \Omega, t \in \mathbb{R}_+,$$

where Z is a compound Poisson process. Solve this equation for X.

3.87 *Piecewise deterministic processes.* In 3.1, let $b = 0$, and let $a = u_0$ satisfy the Lipschitz condition 2.4, and j satisfy 3.2b. Then, between two consecutive jumps, the path $t \mapsto X_t(\omega)$ satisfies the ordinary differential equation

$$\frac{d}{dt} x_t = a(x_t),$$

whose solution is unique and deterministic given its initial condition. Show that Theorem 3.27 holds with the generator

$$Gf(x) = \sum_{i=1}^d a^i(x)\ \partial_i f(x) + \int_0^\infty \ dv\, [f(x + j(x,v)) - f(x)]\,, \qquad f \in \mathcal{C}_K^1.$$

3.88 *Probably no jumps.* This is to give an example of X that has at most finitely many jumps. Suppose that $E = \mathbb{R}$,

$$a(x) = -1, \qquad b(x) = 1, \qquad j(x,v) = -1_{\mathbb{R}_+}(x)1_{(0,c)}(v).$$

Note that $\bar{X}$ of 3.3 is a Brownian motion with downward drift. Show (see Chapter V for this) that

$$A = \text{Leb}\left\{t \geq 0 : \bar{X}_t \in \mathbb{R}_+\right\}$$

is almost surely finite. Show that, for the time R of first jump,

$$\mathbb{P}^x\left\{R = +\infty | \bar{X}\right\} = e^{-cA}.$$

Conclude that $\mathbb{P}^x\{R = +\infty\} = \mathbb{E}^x e^{-cA}$ is strictly positive.

3.89 *Brownian motion plus jumps.* Let X be as in 3.1–3.2. Suppose that $E = \mathbb{R}$ and $\bar{X} = X_0 + W$, a standard Brownian motion. Define $j(x,v) = -ax$ for $0 \leq v \leq 1 - e^{-|x|}$, and $j(x,v) = 0$ otherwise, where a is a constant in (0,1).

a) Describe the motion X during $[0,R)$ and at R, where R is the time of first jump.

b) Specify the Lévy kernel L

c) Specify the generator G given by 3.19.

3.90 *Downward staircase.* Let X satisfy 3.58 with $X_0 = x_0 \in (0,1]$ and

$$j(x,v) = x^3 v \; 1_{[0,1]}(x^2 v), \qquad x \in [0,1], \; v \in \mathbb{R}_+.$$

Let $T_1, T_2, \ldots$ be the successive jump times and define $T_\alpha = \lim \; T_n$.

a) Describe the Markov chain (Y_n), where $Y_n = X_{T_n}$.

b) Show that $Y_1, Y_2, \ldots$ are the atoms of a Poisson random measure on the interval $(0, x_0)$. What is the mean measure?

c) Compute $\mathbb{E}^{x_0} \; T_\alpha$. Note that $X_t = 0$ on $\{t \geq T_\alpha\}$.

d) Compute the transition function (P_t) for X.

3.91 *Transition function for 3.65.* Let X be the process described in Example 3.65. Let P_t and P_t^* be as defined by 3.68. In view of Proposition 3.71, we assume that (P_t^*) is known. This is to compute (P_t). We use $\mathbb{P}^\mu = \int_D \mu(dx)\mathbb{P}^x$.

a) Show that $\mathbb{P}^x\{T_\alpha \leq t\} = 1 - P_t^*(x, \; D)$. Thus,

$$\nu(B) = \mathbb{P}^\mu\left\{T_\alpha \in B\right\} = \int_D \mu(dx) \; \mathbb{P}^x\left\{T_\alpha \in B\right\}, \qquad B \in \mathcal{B}_{\mathbb{R}_+},$$

is well-specified. Let ν^n be the n-fold convolution of ν with itself, with $\nu^0 = \delta_0$ obviously.

b) Define $\rho = \sum_{n=0}^{\infty} \nu^n$. Obviously,

$$\rho(B) = \mathbb{E}^\mu \sum_{n=0}^{\infty} I(T_{n\alpha}, B).$$

Show that $\rho(B) < \infty$ for B compact. Hint:

$$\mathbb{P}^\mu \{T_{n\alpha} \le t\} = \mathbb{P}^\mu \left\{e^{-Tn\alpha} \ge e^{-t}\right\} \le e^t \, \mathbb{E}^\mu e^{-T_{n\alpha}} = e^t \, (\mathbb{E}^\mu \, e^{-T_\alpha})^n.$$

c) Show that $\mathbb{E}^\mu f \circ X_t = \int_{[0,t]} \rho(ds) \int_D \mu(dx) P^*_{t-s} f(x)$.

d) Show that $P_t f(x) = P_t{}^* f(x) + \int_{[0,t]} \nu(ds) \mathbb{E}^\mu f \circ X_{t-s}$.

3.92 *Step processes with discrete state spaces.* Let D be a countable set; we identify it with $\mathbb{N}$ or a subset of $\mathbb{N}$, and regard D as a subset of $E = \mathbb{R}$. We use the notational principles mentioned in Exercise 1.35.

Let X be a step process (right-continuous) satisfying 3.58 and whose values are in D. Then, its Lévy kernel satisfies

$$L(x, A) = \sum_{y \in A} \ell(x, y), \qquad x \in D, A \subset D$$

for some positive numbers $\ell(x, y)$ with

$$\ell(x, x) = 0, \qquad k(x) = \sum_{y \in D} \ell(x, y) < \infty.$$

We may assume that the jump function j has the following form: For each x, let $\{A_{xy} : y \in D\}$ be a partition of $[0, k(x)]$ such that, for each y, the set A_{xy} is an interval of length $\ell(x, y)$. Then, put

$$j(x, v) = \sum_{y \in D} (y - x) 1_{A_{xy}}(v), \qquad x \in D, v \in \mathbb{R}_+.$$

Show that the generator G of X has the form

$$Gf(x) = \sum_{y \in D} \; g(x, y) f(y), \qquad x \in D,$$

and identify the entries $g(x, y)$ of the matrix G. Let $p_t(x, y) = \mathbb{P}^x\{X_t = y\}$ as before in 1.35. show that

$$\frac{d}{dt} \, p_t(x, y) = \sum_{z \in D} p_t(x, z) \; g(z, y), \qquad x, y \in D,$$

and also

$$\frac{d}{dt} \, p_t(x, y) = \sum_{z \in D} g(x, z) \; p_t(z, y), \qquad x, y \in D.$$

If the kernel L is bounded, that is, if the function k is bounded, then we have

$$P_t = e^{tG} = \sum_{n=0}^{\infty} \frac{t^n}{n!} G^n,$$

where P_t is the matrix whose (x,y)-entry is $p_t(x, y)$, and G is the matrix with entries $g(x, y)$ similarly.

3.93 *Semigroups on discrete spaces.* Let D be a countable set. Let the matrices $P_t = [p_t(x, y)]$ satisfy $P_t P_u = P_{t+u}$. Without reference to a Markov process, suppose that $\lim_{t\to 0} p_t(x, x) = 1$ for every x in D. Then, it can be shown that

$$g(x, y) = \lim_{t\to 0} \frac{d}{dt}\, p_t(x, y), \qquad x, y \in D,$$

exist and satisfy $g(x, x) = -k(x)$, with $k(x) \in [0, +\infty]$, and $g(x, y) \in [0, \infty)$ for $x \neq y$, and $\sum\limits_{y\neq x} g(x, y) \le k(x)$. The state x is a trap if $k(x) = 0$, is holding (stable is another term for the same) if $0 < k(x) < \infty$, and is instantaneous if $k(x) = +\infty$. (P_t) is said to be conservative $\sum_{y\neq x} g(x, y) = k(x)$ for every x.

3.94 *Continuation.* Let $D = \mathbb{N}$. For x and y in D, let

$$p_t(x, y) = \mathbb{P}\left\{Y_{s+t} = y | Y_s = x\right\},$$

where Y is the process defined by 1.33. Show that $p_t(x, x) \to 1$ as $t \to 0$. Show that $k(x) \in (0, \infty)$ for each x; identify $k(x)$ in terms of the data $m(q)$, rational q. Show that

$$g(x, y) = 0, \qquad x \neq y.$$

3.95 *Itô processes.* These are Markov processes X that satisfy a stochastic integral equation of the form

$$\begin{aligned} X_t = X_0 + \int_0^t a \circ X_s \, ds + \int_0^t \, b \circ X_s \, dW_s \\ + \int_{[0,t]\times\mathbb{R}_+} (M(ds, dv) - ds\, dv) j(X_{s-}, v) 1_{\{j(X_{s-},v)\le 1\}} \\ + \int_{[0,t]\times\mathbb{R}_+} M(ds, dv) j(X_{s-}, v) 1_{\{j(X_{s-},v)>1\}}. \end{aligned}$$

Here, a, b, M, W, j are as in 3.1, but without the condition 3.2, and j must satisfy

$$\int_{\mathbb{R}_+} du \left[\left(j(x,\, v)^2 \wedge 1\right)\right] < \infty,$$

and the third integral is a stochastic integral, defined as a limit in probability. This class of processes includes all Lévy processes (see Itô-Lévy decomposition), all Itô diffusions, all jump-diffusions, and more. See the complement 5.51 for more.

4 Markov Systems

This section is to introduce Markov processes in the modern setting. We shall introduce a probability measure $\mathbb{P}^x$ for each state x; it will serve as the conditional probability law given that the process X is at x initially. We shall introduce a shift operator θ_t for each time t; it will indicate that t is the present time. And, we shall think of X as the motion of a particle that lives in E, but might die or be killed at some random time; this will require an extra point to serve as the cemetery.

The space E will be kept fairly general. Although the Markov property has nothing to do with the topology of E, the analytical machinery requires that E be topological and X right-continuous. The reader is invited to take $E = \mathbb{R}$ on a first reading. This section is independent of Sections 2 and 3; but some familiarity with at least Section 2 would be helpful as motivation. Also helpful is the formalism of Lévy processes; the connections are spelled out in Exercises 4.31 and 4.32.

The system

This is to describe the setting for Markov processes. The time-set is $\mathbb{R}_+$; it will be extended to $\bar{\mathbb{R}}_+$.

4.1 *State space.* Let E be a locally compact separable metrizable space, and $\mathcal{E}$ the Borel σ-algebra on it. If E is compact, we let ∂ be an isolated point outside E. If E is not compact, ∂ will be the "point at infinity" in the one point compactification of E. We put

$$\bar{E} = E \cap \{\partial\}, \qquad \bar{\mathcal{E}} = \sigma(\mathcal{E} \cup \{\bar{E}\}).$$

4.2 *Convention.* Every function $f: E \mapsto \bar{\mathbb{R}}$ is extended onto $\bar{E}$ automatically by setting $f(\partial) = 0$. Thus, writing $f \in \mathcal{E}$ indicates also a function in $\bar{\mathcal{E}}$ with $f(\partial) = 0$; otherwise, we write $\bar{f} \in \bar{\mathcal{E}}$ to mean that $\bar{f}$ is defined on $\bar{E}$ and is $\bar{\mathcal{E}}$-measurable without an assumption on $\bar{f}(\partial)$.

4.3 *Transition semigroups.* Let (P_t) be a family of sub-Markov kernels on $(E, \mathcal{E})$ such that $P_t P_u = P_{t+u}$. Each P_t is extended to become a Markov kernel $\bar{P}_t$ on $(\bar{E}, \bar{\mathcal{E}})$ by putting

$$\bar{P}_t(x, B) = P_t(x, B \cap E) + (1 - P_t(x, E))I(\partial, B), \qquad x \in \bar{E}, B \in \bar{\mathcal{E}}.$$

Note that $\bar{P}_t(\partial, B) = I(\partial, B) = 1_B(\partial)$ by the preceding convention applied to the function $x \mapsto P_t(x, E)$ on E, namely, the convention that puts $P_t(\partial, E) = 0$. It is easy to check that $\bar{P}_t \bar{P}_u = \bar{P}_{t+u}$.

4.4 *Stochastic base.* Let $(\Omega, \mathcal{H})$ be a measurable space, $\mathcal{F} = (\mathcal{F}_t)$ a filtration over it, and $\theta = (\theta_t)$ a family of "shift" operators $\theta_t : \Omega \mapsto \Omega$ such that $\theta_0 \omega = \omega$ and

$$\theta_u(\theta_t \omega) = \theta_{t+u}\, \omega$$

for every ω in Ω. We assume that there is a special point ω_∂ in Ω, and that $\theta_t\omega_\partial = \omega_\partial$ for all t, and $\theta_\infty\omega = \omega_\partial$ for all ω. Finally, let $\mathbb{P}^\bullet = (\mathbb{P}^x)$ be a family of probability measures $\mathbb{P}^x$ on $(\Omega, \mathcal{H})$ such that $(x, H) \mapsto \mathbb{P}^x(H)$ is a transition kernel from $(\bar{E}, \bar{\mathcal{E}})$ into $(\Omega, \mathcal{H})$.

4.5 *Stochastic process.* Let $X = (X_t)$ be a process with state space $(\bar{E}, \bar{\mathcal{E}})$, adapted to the filtration $\mathcal{F}$, and with the point ∂ as a trap; the last phrase means that if $X_t(\omega) = \partial$ then $X_{t+u}(\omega) = \partial$ for all $u \geq 0$. We assume that $X_0(\omega_\partial) = \partial$, and that $X_\infty(\omega) = \partial$ for all ω, and that

$$X_u(\theta_t\omega) = X_{t+u}(\omega), \qquad \omega \in \Omega, \quad t, u \in \bar{\mathbb{R}}_+.$$

We let $\mathcal{G}^o = (\mathcal{G}^o_t)$ be the filtration generated by X, and put $\mathcal{G}^o_\infty = \vee_t\ \mathcal{G}^o_t$ as usual.

Markov system

Throughout this section and further we are working with the system described in 4.1–4.5 above. In conditional expectations and probabilities, we use the old conventions (see V.2.21 *et seq.*) and put

4.6 $$\mathbb{P}^x_T = \mathbb{P}^x(\cdot|\mathcal{F}_T), \qquad \mathbb{E}^x_T = \mathbb{E}^x(\cdot|\mathcal{F}_T).$$

The following is the enhanced version of Markovness.

4.7 DEFINITION. *The system* $\mathfrak{X} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X, \mathbb{P}^\bullet)$ *is said to be* Markov *with* living space E *and transition semigroup* (P_t) *if the following hold:*

Normality. $\mathbb{P}^x\{X_0 = x\} = 1$ *for every* x *in* $\bar{E}$.

Right-continuity for $\mathcal{F}$. *The filtration* $(\mathcal{F}_t)$ *is right-continuous.*

Regularity of paths. *For every* ω, *the path* $t \mapsto X_t(\omega)$ *is right-continuous and has left-limits as a function from* $\mathbb{R}_+$ *into* $\bar{E}$.

Markov property. *For every* x *in* E *and every* t *and* u *in* $\mathbb{R}_+$,

4.8 $$\mathbb{E}^x_t\ f \circ X_{t+u} = P_u f \circ X_t, \qquad f \in \mathcal{E}_+. \qquad \square$$

The normality condition makes $\mathbb{P}^x$ the probability measure on $(\Omega, \mathcal{H})$ under which X is started at x. The right-continuity of $\mathcal{F}$ enriches the pool of stopping times and will be of further use with the strong Markov property; note that $\mathcal{G}^o_{t+} \subset \mathcal{F}_{t+} = \mathcal{F}_t$.

4.9 REMARK. In terms of the definitions of Section 1, the Markov property of the preceding definition implies the following for each x in $\bar{E}$: Over the probability space $(\Omega, \mathcal{H}, \mathbb{P}^x)$, the process X is a (time-homogeneous) Markov process with state space $(\bar{E}, \bar{\mathcal{E}})$ and transition function $(\bar{P}_t)$ given in 4.3. This can be seen by noting that, in view of the conventions, 4.8 implies that

$$\mathbb{E}^x\ \bar{f} \circ X_{t+u} = \bar{P}_u \bar{f} \circ X_t, \qquad \bar{f} \in \bar{\mathcal{E}}_+.$$

Thus, Definition 4.7 introduces a family of Markov processes, one for each x in $\bar{E}$, but all these processes have the same transition function and are intertwined in a systematic manner.

4.10 REMARK. The meaning of P_t is implicit in 4.8. There, putting $t = 0$, applying the expectation operator $\mathbb{E}^x$ to both sides, and using the normality,

4.11 $$\mathbb{E}^x\ f \circ X_u = P_u f(x), \qquad x \in E, u \in \mathbb{R}_+,\ f \in \mathcal{E}_+.$$

This remains true for $x = \partial$ as well, because $X_u = \partial$ almost surely under $\mathbb{P}^\partial$, and the conventions yield $f(\partial) = 0$ and $P_u f(\partial) = 0$. Hence, we may re-write the condition 4.8, using $X_{t+u} = X_u \circ \theta_t$ as in 4.5, in the form

4.12 $$\mathbb{E}_t^x\ f \circ X_u \circ \theta_t = \mathbb{E}^{X_t}\ f \circ X_u, \qquad f \in \mathcal{E}_+,$$

but with a cautionary remark: the right side stands for $g \circ X_t$ where

4.13 $$g(y) = \mathbb{E}^y f \circ X_u = P_u f(y).$$

Almost surely

Since there is a multitude of probability measures $\mathbb{P}^x$, it is convenient to say "almost surely" to mean "almost surely under $\mathbb{P}^x$ for every x in $\bar{E}$". Similarly, a proposition $\pi(\omega)$ for ω is said to hold for almost every ω in H, and then we write

4.14 $$\pi \text{ } a.e. \text{ } on \text{ } H, \text{ or, } \pi(\omega) \quad \text{for } a.e.\ \omega \text{ in } H,$$

if $H \in \mathcal{H}$ and for every x in $\bar{E}$ there is a subset H_x of H in $\mathcal{H}$ such that $\mathbb{P}^x(H \backslash H_x) = 0$ and $\pi(\omega)$ holds for every ω in H_x.

Lifetime of X

According to 4.5, the "boundary" point ∂ is a trap; it is the final resting place for X. Thus

4.15 $$\zeta = \inf\{t \in \mathbb{R}_+ :\ X_t = \partial\}$$

is called the *lifetime* of X. Note that $\zeta(\omega) > u$ if and only if $X_u(\omega) \in E$; hence the term "living space" for E. When $\mathfrak{X}$ is Markov, it follows from 4.11 with $f = 1_E$ that

$$\mathbb{P}^x\{\zeta > u\} = P_u(x, E), \qquad x \in E, \qquad u \in \mathbb{R}_+.$$

This gives meaning to the defect $1 - P_u(x,\ E)$ when P_u is sub-Markov. The process X is said to be *conservative* if $P_t(x,\ E) = 1$ for all x in E, that is, if every P_t is Markov.

Markov property

Here we explore the essential condition in Definition 4.7, the Markov property.

4.16 REMARK. The collection of functions f in $\mathcal{E}$ for which 4.8 holds is a monotone class. Thus, in order for 4.8 to hold for all f in $\mathcal{E}_+$, it is sufficient that it hold for the indicators of Borel subsets of E, or for the indicators of open subsets of E, or for bounded continuous functions on E, or for continuous functions on E with compact support.

The next theorem captures the essence of Markov property by replacing $f \circ X_u$ in 4.12 with an arbitrary functional V of X. Note, for example, that

$$V = f(X_{u_1}, \ldots, X_{u_n}) \Rightarrow V \circ \theta_t = f(X_{t+u_1}, \ldots, X_{t+u_n}).$$

Thus, the proof of the next theorem is immediate from Remark 4.10, Theorem 1.2, and 4.12 above.

4.17 THEOREM. (Markov property). *Suppose that $\mathfrak{X}$ is a Markov system. Then, for every x in $\bar{E}$ and t in $\mathbb{R}_+$ and positive V in $\mathcal{G}^o_\infty$,*

$$\mathbb{E}^x_t \; V \circ \theta_t = \mathbb{E}^{X_t} V.$$

4.18 EXAMPLE. This is to illustrate the preceding theorem with a specific V. The aim is to clarify some technical matters which are implicit, and also to re-iterate the heuristics.

a) Let $f \in \mathcal{E}_b$, that is, let $f : E \mapsto \mathbb{R}$ be a bounded Borel function. Since X is right-continuous and each X_t is measurable with respect to $\mathcal{G}^o_\infty$ and $\bar{\mathcal{E}}$, the mapping $(t, \omega) \mapsto X_t(\omega)$ is measurable with respect to $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{G}^o_\infty$ and $\bar{\mathcal{E}}$. Thus, $(t, \omega) \mapsto f \circ X_t(\omega)$ is in $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{G}^o_\infty$.

b) Hence, Fubini's theorem shows that, for fixed $p > 0$,

$$V = \int_{\mathbb{R}_+} du \; e^{-pu} \; f \circ X_u$$

defines a bounded variable V in $\mathcal{G}^o_\infty$; and

$$g(y) = \mathbb{E}^y \; V, \qquad y \in E,$$

defines a function g in $\mathcal{E}_b$; this is because $\mathbb{P}^\bullet$ is a transition kernel, and $g(y) = \mathbb{P}^y V$ in the kernel-function notation. Similarly, and since $X_u \circ \theta_t = X_{t+u}$ by 4.5,

$$V \circ \theta_t = \int_{\mathbb{R}_+} du \; e^{-pu} \; f \circ X_{t+u}$$

is a well-defined bounded random variable in $\mathcal{G}^o_\infty$.

c) Now, the heuristic part: $g(y)$ is our estimate of V made at time 0 if the initial state is y. The initial state of the process $X \circ \theta_t$ is X_t. The variable $V \circ \theta_t$ is the same functional of $X \circ \theta_t$ as V is of X. Thus, our estimate of $V \circ \theta_t$ made at time t should be $g(X_t)$ if we think that t is the origin of time and all the past is past. □

Blumenthal's zero-one law

This useful result is a consequence of the normality, the Markov property, and the right-continuity of the filtration $\mathcal{F}$.

4.19 THEOREM. *Let $\mathfrak{X}$ be Markov. Let H be an event in $\mathcal{G}^o_{0+}$. For each x in $\bar{E}$, then, $\mathbb{P}^x(H)$ is either 0 or 1.*

Proof. Put $V = 1_H$. Clearly, $V = V \cdot V$, and $V = V \circ \theta_0$ since $\theta_0\omega = \omega$ for all ω; hence, $V = V \cdot (V \circ \theta_0)$. On the other hand, the filtration $(\mathcal{G}^o_t)$ generated by X is coarser than $(\mathcal{F}_t)$ by the adaptedness of X to $\mathcal{F}$; and, thus, $\mathcal{G}^o_{t+} \subset \mathcal{F}_{t+} = \mathcal{F}_t$, the last equality being the definition of right-continuity for $\mathcal{F}$. This implies, since $V \in \mathcal{G}^o_{0+}$ by assumption, that $V \in \mathcal{F}_0$. It now follows from the Markov property 4.17 at $t = 0$ that, since $\mathbb{E}^x = \mathbb{E}^x\mathbb{E}^x_0$,

$$\mathbb{E}^x\ V = \mathbb{E}^x V \cdot V \circ \theta_0 = \mathbb{E}^x V \mathbb{E}^{X_0} V.$$

But, by normality, $X_0 = x$ with $\mathbb{P}^x$-probability one. Hence, $\mathbb{E}^x V = \mathbb{E}^x V\ \mathbb{E}^x V$, which implies that $\mathbb{E}^x V = \mathbb{P}^x(H)$ is either 0 or 1. □

Holding points, instantaneous points

Started at a point x, the process either exits x instantaneously or stays at x some strictly positive amount of time. This dichotomy is a consequence of the preceding zero-one law.

Suppose that the system $\mathfrak{X}$ is Markov. Define

4.20 $$R = \inf \{t > 0 : X_t \neq X_0\}.$$

Then, R is a stopping time of $\left(\mathcal{G}^o_{t+}\right)$, and thus, the event $\{R = 0\}$ belongs to $\mathcal{G}^o_{0+}$. Hence, for fixed x in $\bar{E}$, the zero-one law applies to show that

4.21 $$\mathbb{P}^x \{R = 0\}$$

is either 0 or 1. It this probability is 0, then x is said to be a *holding point*; if it is 1, then x is said to be *instantaneous.* A holding point x is called a *trap*, or an *absorbing* point, if $\mathbb{P}^x\{R = \infty\} = 1$.

The point ∂ is a trap; there may be other traps. For step processes of Section 3, and for Poisson and compound Poisson processes, every point of E is a holding point. For Brownian motions in $E = \mathbb{R}^d$, every point of E is instantaneous; similarly for Itô diffusions. For Lévy processes other than compound Poisson, every point of $E = \mathbb{R}^d$ is instantaneous.

Let x be a holding point. Started at x, the process stays there for a strictly positive amount R of time. The next theorem shows that the distribution of R is exponential, and the state X_R is independent of R. We shall show later (see 5.23) that when X is strong Markov, it must exist x by a jump.

4.22 THEOREM. *Let $\mathfrak{X}$ be Markov, and let x be a holding point. Then,*

$$\mathbb{P}^x\{R > t, X_R \in B\} = e^{-k(x)t}K(x,B),\ t \in \mathbb{R}_+,\ B \in \bar{\mathcal{E}},$$

for some number $k(x) < \infty$ and some measure $B \mapsto K(x,B)$ on $(\bar{E}, \bar{\mathcal{E}})$.

Proof. For every ω, the definition of R implies that

$$R(\omega) > t + u \Leftrightarrow R(\omega) > t \quad \text{and} \quad R(\theta_t\omega) > u,$$

and, then,

$$X_R(\omega) = X_{t+R(\theta_t\omega)}(\omega) = X_{R(\theta_t\omega)}(\theta_t\omega) = X_R(\theta_t\omega).$$

Thus, for B in $\bar{\mathcal{E}}$,

4.23
$$\begin{aligned}\mathbb{P}^x\{R > t+u,\ X_R \in B\} &= \mathbb{P}^x\{R > t,\ R \circ \theta_t > u,\ X_R \circ \theta_t \in B\}\\ &= \mathbb{E}^x 1_{\{R>t\}} \mathbb{P}^{X_t}\{R > u, X_R \in B\}\\ &= \mathbb{P}^x\{R > t\}\, \mathbb{P}^x\{R > u, X_R \in B\};\end{aligned}$$

here, the second equality is justified by the Markov property of 4.17 at time t, and the third equality by the observation that $X_t = X_0$ on $\{R > t\}$ followed by the normality condition.

In 4.23, take $B = \bar{E}$. The result is

$$\mathbb{P}^x\{R > t+u\} = \mathbb{P}^x\{R > t\}\, \mathbb{P}^x\{R > u\}$$

for all t and u in $\mathbb{R}_+$; and $t \mapsto \mathbb{P}^x\{R > t\}$ is obviously right-continuous and is equal to 1 at $t = 0$ (since x is holding). Thus, there exists $k(x)$ in $\mathbb{R}_+$ such that

4.24
$$\mathbb{P}^x\{R > t\} = e^{-k(x)t}, \qquad t \in \mathbb{R}_+.$$

Next, put this into 4.23 and set $u = 0$. Since x is holding, $\mathbb{P}^x\{R > 0\} = 1$, and 4.23 becomes

$$\mathbb{P}^x\{R > t, X_R \in B\} = e^{-k(x)t}\ \mathbb{P}^x\{X_R \in B\},$$

which has the form claimed. □

4.25 REMARK. The point x is a trap if and only if $k(x) = 0$. In fact, 4.24 holds for instantaneous x as well; then, $k(x) = +\infty$.

Measures P^μ

For each x in $\bar{E}$, the distribution of X_0 under $\mathbb{P}^x$ is the Dirac measure δ_x; this is by the normality of $\mathfrak{X}$. Thus, for an arbitrary probability measure on $(\bar{E}, \bar{\mathcal{E}})$,

4.26
$$\mathbb{P}^\mu(H) = \int_E \mu(dx)\, \mathbb{P}^x(H), \qquad H \in \mathcal{H},$$

defines a probability measure on $(\Omega, \mathcal{H})$, under which X_0 has the distribution μ. This follows from Theorem I.6.3 via the hypothesis in 4.4 that $(x,\ H) \mapsto \mathbb{P}^x(H)$ is a transition probability kernel.

Exercises

4.27 *Compound Poissons.* Let $X = X_0 + Y$, where Y is a compound Poisson process (with $Y_0 = 0$) whose jump times form a Poisson process with rate c, and whose jump sizes have the distribution μ. Classify the states as holding or instantaneous. What are $k(x)$ and $K(x, B)$ of Theorem 4.22 in this case?

4.28 *Step processes.* Let X be a step process as in Section 3. Show that every point x in $E = \mathbb{R}^d$ is a holding point. Compute $k(x)$ and $K(x, B)$ of Theorem 4.22 in terms of the Lévy kernel L of X.

4.29 *Brownian motion with holding boundary points.* Let $a < 0 < b$ be fixed. Started in the interval $(a,\ b)$, the motion X is standard Brownian until it exits the interval; if the exit is at a, then X stays at a an exponential time with parameter $k(a)$ and then jumps to the point 0; if the exit is at b, then X stays at b an exponential time with parameter $k(b)$ and then jumps to 0. Once at 0, the motion resumes its Brownian character, and so on. Classify the points of $E = [a, b]$; identify the distributions $K(x, \cdot)$ for the holding points x.

4.30 *Achilles' run.* The living space E is (0,1]; then $\partial = 0$ necessarily. Started at x in E, the particle stays at x an exponential amount of time with mean x and, then, jumps to $y = x/2$; at y, it stays an exponential time with mean y and, then, jumps to $z = y/2$; and so on. Let $T_1,\ T_2, \ldots$ be the successive jump times, put $\zeta = \lim\ T_n$, and define $X_t = \partial = 0$ for t in $[\zeta, +\infty]$. Show that all points are holding points. Identify the parameters $k(x)$ and $K(x, B)$. Compute $\mathbb{E}^x\zeta$.

4.31 *Lévy processes.* Let $\mathfrak{X} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X, \mathbb{P}^\bullet)$ be a Markov system with living space $E = \mathbb{R}^d$. Suppose that its transition semigroup (P_t) is such that, for each t,

$$P_t f(x) = \int_E \pi_t(dy)\ f(x + y), \qquad x \in E,\ f \in \mathcal{E}_+.$$

for some probability measure π_t on E. Show that, then, X has stationary and independent increments under each $\mathbb{P}^x$. More precisely, for each x in E, the process $Y = (X_t - X_0)_{t \in \mathbb{R}_+}$ is a Lévy process over the stochastic base $(\Omega, \mathcal{H}, \mathcal{F}^x, \theta, \mathbb{P}^x)$ in the sense of Definition VII.3.3; here $\mathcal{F}^x$ is the augmentation of $\mathcal{F}$ with respect to the probability measure $\mathbb{P}^x$. Show this.

4.32 *Continuation.* This is a converse to the preceding. Let X and $\mathcal{B} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$ be as in Definitions VII.3.1 and VII.3.3. Put $E = \mathbb{R}^d,\ \mathcal{E} = \mathcal{B}(\mathbb{R}^d)$, and set ∂ to be the "point at infinity." Define

$$\hat{\Omega} = \bar{E} \times \Omega,\ \hat{\mathcal{H}} = \bar{\mathcal{E}} \otimes \mathcal{H},\ \hat{\mathcal{F}}_t = \bar{\mathcal{E}} \otimes \mathcal{F}_t, \hat{\mathbb{P}}^x = \delta_x \times \mathbb{P}$$

for x in $\bar{E}$; and, for $\hat{\omega} = (x, \omega)$ in $\hat{\Omega}$, put

$$\hat{X}_t(\hat{\omega}) = x + X_t(\omega), \qquad \hat{\theta}_t \hat{\omega} = (\hat{X}_t(\hat{\omega}), \theta_t \omega).$$

Show that $\hat{\mathfrak{X}} = (\hat{\Omega}, \hat{\mathcal{H}}, \hat{\mathcal{F}}, \hat{\theta}, \hat{X}, \hat{\mathbb{P}}^\bullet)$ is a Markov system, in the sense of Definition 4.7, with living space E and transition function (P_t) given by

$$P_t(x, B) = \mathbb{P}\{x + X_t \in B\}.$$

5 Hunt Processes

These are Markov processes which have almost all the properties desired of a Markov process. Itô diffusions, jump-diffusions, simple step processes, and all Lévy processes (including, of course, Poisson and Brownian motions) are Hunt processes. We choose them to form the central reference system for the theory; even when a Markov process is not Hunt, it is best to describe it by telling how it differs from a Hunt process.

Throughout this section, $\mathfrak{X} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X, \mathbb{P}^\bullet)$ is a Markov system with living space E and transition semigroup (P_t); see Definition 4.7 and the setup 4.1–4.6. Recall, in particular, that the filtration $\mathcal{F}$ is right-continuous and that the path $t \mapsto X_t$ is right-continuous and has left-limits in $\bar{E}$. Recall also that $(\mathcal{G}_t^o)$ is the filtration generated by X.

In preparation for the definition of strong Markov property next, we note that $\bar{f} \circ X_T$ is $\mathcal{F}_T$-measurable for every $\mathcal{F}$-stopping time T and every $\bar{\mathcal{E}}$-measurable $\bar{f} : \bar{E} \mapsto \bar{\mathbb{R}}_+$. For continuous $\bar{f}$, this follows from Theorem V.1.14 via the right-continuity of $\bar{f} \circ X$. Then, a monotone class argument extends it to all $\bar{\mathcal{E}}$-measurable $\bar{f}$. For positive V in $\mathcal{G}_\infty^o$, putting $\bar{f}(y) = \mathbb{E}^y V$ yields a function $\bar{f}$ that is $\bar{\mathcal{E}}$-measurable; and, then, $\bar{f} \circ X_T$ belongs to $\mathcal{F}_T$ as required for it to be a conditional expectation given $\mathcal{F}_T$.

5.1 Definition. *The Markov system $\mathfrak{X}$ is said to be* strong Markov *if, for every $\mathcal{F}$-stopping time T and every positive random variable V in $\mathcal{G}_\infty^o$,*

5.2 $$\mathbb{E}_T^x \; V \circ \theta_T = \mathbb{E}^{X_T} \; V, \qquad x \in E.$$

It is said to be quasi-left-continuous if, for every increasing sequence (T_n) of $\mathcal{F}$-stopping times with limit T,

5.3 $$\lim_n \; X_{T_n} = X_T \text{ almost surely on } \{\mathrm{T} < \infty\}.$$

It is said to be a Hunt system *if it is strong Markov and quasi-left-continuous.*

We shall explore the contents of these definitions and their ramifications. We start with the less familiar concept.

Quasi-left-continuity

If X is continuous, then $\mathfrak{X}$ is quasi-left-continuous automatically. The continuity is not necessary. For instance, if $X - X_0$ is a Poisson process, then $\mathfrak{X}$ is quasi-left-continuous even though X has infinitely many jumps.

Similarly, if $X - X_0$ is a Lévy process, or if X is a jump-diffusion or a step process, then $\mathfrak{X}$ is quasi-left-continuous. These comments will become clear shortly.

Recall the notation X_{t-} for the left-limit of X at t; we put $X_{0-} = X_0$ for convenience. For a random time T, then, X_{T-} is the random variable $\omega \mapsto X_{T-}(\omega) = X_{T(\omega)-}(\omega)$. Suppose now that (T_n) is an increasing sequence of $\mathcal{F}$-stopping times with limit T, and pick ω such that $T(\omega) < \infty$. If $T_n(\omega) = T(\omega)$ for all n large enough, then $\lim X_{T_n}(\omega) = X_T(\omega)$ trivially. Otherwise, if

5.4 $$T_n(\omega) < T(\omega) \quad \text{for all } n,$$

then $\lim\ X_{T_n}(\omega) = X_{T-}(\omega)$, and quasi-left-continuity would require that $X_{T-}(\omega) = X_T(\omega)$, unless ω happens to be in the negligible exceptional set of 5.3. In other words, if $\mathfrak{X}$ is quasi-left-continuous, then 5.4 is incompatible with

5.5 $$T(\omega) < \infty,\ X_{T-}(\omega) \neq X_T(\omega),$$

except for a negligible set of ω.

We may interpret 5.4 as "predictability" for $T(\omega)$, because the sequence of times $T_n(\omega)$ enables the observer to foresee $T(\omega)$. So, heuristically, quasi-left-continuity is about the continuity of paths at predictable times and, equivalently, about the unpredictability of jump times. We make these remarks precise next.

Predictable times, total unpredictability

We recall some definitions introduced in passing in Chapter V, adapted to the newer meaning of "almost everywhere" given around 4.14. We shall use the notation (read T on H)

5.6 $$T_H(\omega) = \begin{cases} T(\omega) & \text{if } \omega \in H, \\ +\infty & \text{otherwise,} \end{cases}$$

for $\mathcal{F}$-stopping times T and events H in $\mathcal{F}_T$; and, then, T_H is also an $\mathcal{F}$-stopping time.

5.7 Definition. *Let T be an $\mathcal{F}$-stopping time. It is said to be* predictable *if there exists an increasing sequence (T_n) of $\mathcal{F}$-stopping times with limit T such that*

5.8 $$T_n < T \text{ for all } n \qquad \text{a.e.} \quad \text{on } \{0 < T < \infty\}.$$

It is said to be totally unpredictable *if, for every predictable $\mathcal{F}$-stopping time S,*

5.9 $$T = S \quad \textit{almost nowhere on } \{T < \infty\}.$$

For Brownian motion, every hitting time is predictable, and more. For a Poisson process, Proposition VI.5.20 implies that the first jump time is totally unpredictable; see also V.7.31. In Example 3.65, the time T_α is predictable, so are $T_{2\alpha}$, $T_{3\alpha}$, etc. The other times in 3.67 are totally unpredictable. The following enhances the definition.

5.10 Lemma. *Let T be a totally unpredictable $\mathcal{F}$-stopping time. Suppose that (T_n) is an increasing sequence of $\mathcal{F}$-stopping times with limit T on the event $\{T < \infty\}$. Then, 5.4 fails for almost every ω in $\{T < \infty\}$.*

Proof. Let $H_n = \{T_n < T\}$ and $H = \cap_n H_n$. We need to show that

5.11 $$\mathbb{P}^x \left(H \cap \{T < \infty\} \right) = 0$$

for every x. Define $S = T_H$, see 5.6, and define S_n similarly from T_n and H_n; these are all $\mathcal{F}$-stopping times. Since (T_n) is increasing, the sequence (H_n) is shrinking to H, and (S_n) is increasing to S. Moreover, if $\omega \in H$, then $S_n(\omega) = T_n(\omega) < T(\omega) = S(\omega)$. Since $\{0 < S < \infty\} \subset H$, we conclude that S is predictable. It follows from the total unpredictability of T that 5.9 holds; and 5.9 is the same as 5.11 for all x. □

Total unpredictability of jumps

Let T be an $\mathcal{F}$-stopping time. We call T a *time* of *continuity* for X if (recall that $X_{0-} = X_0$)

5.12 $$X_{T-} = X_T \qquad \text{a.e. on } \{T < \infty\} ,$$

and *a jump time* for X if

5.13 $$X_{T-} \neq X_T \qquad \text{a.e. on } \{T < \infty\} .$$

The following clarifies the true meaning of quasi-left-continuity.

5.14 Theorem. *The following are equivalent:*

a) The Markov system $\mathfrak{X}$ is quasi-left-continuous.

b) Every predictable $\mathcal{F}$-stopping time is a time of continuity.

c) Every jump time is totally unpredictable.

Proof. Suppose (a). Let T be predictable. Then, there is (T_n) increasing to T such that 5.8 holds. Therefore, $\lim X_{T_n} = X_{T-}$ a.e. on $\{T < \infty\}$. But the limit is X_T a.e. on $\{T < \infty\}$ by the assumed quasi-left-continuity. Thus, 5.12 holds. Hence $(a) \Rightarrow (b)$.

Suppose (b). Let T be a jump time, that is, let 5.13 hold, and let S be predictable. Then, $X_{T-} = X_{S-} = X_S = X_T$ almost everywhere on $\{T = S,\ T < \infty\}$ in view of (b) for S. This means, in view of 5.13 for T, that 5.9 holds. Hence, T is totally unpredictable. So, $(b) \Rightarrow (c)$.

Suppose (c). Let (T_n) be an increasing sequence of $\mathcal{F}$-stopping times with limit T. On $\{X_{T-} = X_T,\ T < \infty\}$, we have $\lim X_{T_n} = X_T$ obviously. To show quasi-left-continuity at T, we show next that

5.15 $$\lim\ X_{T_n} = X_T \qquad \text{a.e. on } H = \{X_{T-} \neq X_T,\ T < \infty\} .$$

Define $S = T_H$ as in 5.6; it is obviously a jump time. By the assumed (c), then, S is totally unpredictable. Moreover, $\lim T_n = T = S$ on $\{S < \infty\}$, because $\{S < \infty\} = H \subset \{T < \infty\}$. It follows from Lemma 5.10 that, almost surely on $\{S < \infty\} = H$, we have $T_n = S$ for all n large enough. Since $S = T$ on H, we have 5.15. So, $(c) \Rightarrow (a)$. □

Classification of stopping times

A hasty reading of the last theorem might suggest that a stopping time is predictable if and only if it is a continuity time. This is false in general; see Example 5.22. However, it is true provided that we limit ourselves to the stopping times of $\left(\mathcal{G}^o_{t+}\right)$. We offer this without proof. Note that, when $\mathfrak{X} = (\ldots, \mathcal{F}, \ldots)$ is quasi-left-continuous, then so is the Markov system $\mathfrak{X}^0 = \left(\ldots, \left(\mathcal{G}^o_{t+}\right), \ldots\right)$; so, half of the statements next follow from the last theorem.

5.16 THEOREM. *Suppose that $\mathfrak{X}$ is quasi-left-continuous. Consider a stopping time of $\left(\mathcal{G}^o_{t+}\right)$. It is predictable if and only if it is a continuity time for X; it is totally unpredictable if and only if it is a jump time for X.* □

5.17 REMARK. Let S be an $\mathcal{F}$-stopping time. We call it *σ-predictable* if there is a sequence (S_n) of predictable $\mathcal{F}$-stopping times such that, for almost every ω with $S(\omega) < \infty$, we have $S(\omega) = S_n(\omega)$ for some n. An arbitrary $\mathcal{F}$-stopping time R can be written as

5.18 $$R = S \wedge T,$$

where S is σ-predictable and T totally unpredictable. The preceding theorem implies, in particular, that for a quasi-left-continuous system, every σ-predictable stopping time of $\left(\mathcal{G}^o_{t+}\right)$ is necessarily predictable. Hence, every $\left(\mathcal{G}^o_{t+}\right)$-stopping time R has the form 5.18 with S predictable and T totally unpredictable; indeed, in the notation 5.6,

$$S = R_{\{X_{R-} = X_R, R < \infty\}}, \qquad T = R_{\{X_{R-} \neq X_R, R < \infty\}}.$$

Examples

All continuous Markov processes are obviously quasi-left-continuous. So, we concentrate on processes with jumps. The reader will see that quasi-left-continuity at a jump time depends on whether that jump is endogeneous (as in the first example below) or exogeneous (and is caused by kicks from a Poisson).

5.19 *Brownian motion with jump boundaries.* This is a variation on Exercise 4.29. The motion X is Brownian inside the interval (a,b) until the time T of exit; if X_{T-} is a, then X_T has some distribution μ_a on (a,b); if X_{T-} is b, then X_T has some distribution μ_b on (a,b). This $\mathfrak{X}$ is strong Markov; it is not quasi-left-continuous. To see the latter point, let T_n be the time of exit

from $(a + 1/n,\ b - 1/n)$; then, (T_n) increases to T, but $X_{T-} \neq X_T$. Note that, in this example, the jumps are triggered by the particle itself. The process of Example 4.29, by contrast, is quasi-left-continuous; its jumps are exogeneous; they are caused by kicks from a Poisson.

5.20 *Step processes.* Suppose that X is a step process. Then $\mathfrak{X}$ is strong Markov, we show now that it is quasi-left-continuous. So, $\mathfrak{X}$ is Hunt.

We start by showing that quasi-left-continuity hold at the time R of first jump. Let (R_n) be a sequence of stopping times increasing to R. Fix x in E and recall that, under $\mathbb{P}^x$, the time R has the exponential distribution with some parameter $k(x) < \infty$.

If x is a trap, then $k(x) = 0$ and $\mathbb{P}^x\{R = +\infty\} = 1$, and thus the condition 5.3 holds at R by default. Suppose that x is not a trap. Observe that, on $\{R_n < R\}$, we have $R = R_n + R \circ \theta_{R_n}$. Thus,

$$\begin{aligned}\mathbb{E}^x R &= \mathbb{E}^x\ R 1_{\{R_n = R\}} + \mathbb{E}^x\left(R_n + R \circ \theta_{R_n}\right)\ 1_{\{R_n < R\}} \\ &= \mathbb{E}^x\ R_n 1_{\{R_n = R\}} + \mathbb{E}^x\ R_n 1_{\{R_n < R\}} + \mathbb{E}^x 1_{\{R_n < R\}} \mathbb{E}^{X(R_n)} R \\ &= \mathbb{E}^x R_n + \mathbb{E}^x R\ \mathbb{P}^x\left\{R_n < R\right\};\end{aligned}$$

here, we used the strong Markov properly at R_n and noted that $X_{R_n} = x$ on $\{R_n < R,\ X_0 = x\}$. Since (R_n) is increasing to R and $\mathbb{E}^x R < \infty$ (since x is not a trap), we conclude that

$$\lim_n \mathbb{P}^x\left\{R_n < R\right\} = 0.$$

But, the events $\{R_n < R\}$ are shrinking to $\{R_n < R \text{ for all } n\}$. So,

$$\mathbb{P}^x\left\{R_n < R \text{ for all n}\right\} = 0,$$

that is, for $\mathbb{P}^x$-almost every ω, we have $R_n(\omega) = R(\omega)$ for all n large enough, and hence, $\lim X_{R_n}(\omega) = X_R(\omega)$. Thus, quasi-left-continuity holds at R.

Fix m in $\mathbb{N}$, let T be the $(m+1)^{th}$ jump time and let (T_n) an increasing sequence of stopping times with limit T. Let S denote the mth jump time and put $R_n = S \vee T_n$. Then, (R_n) is increasing to T, and $S \le R_n \le T$, and $T = S + R \circ \theta_S$ with R as before (the time of first jump). The arguments of the last paragraph apply with the conditional law $\mathbb{P}^x_S$ replacing $\mathbb{P}^x$; this is by the strong Markov property at S. Thus, almost surely on $\{T < \infty\},\ R_n = T$ for all n large enough and

$$\lim_n X_{T_n} = \lim X_{R_n} = X_T.$$

So, quasi-left-continuity holds at T; and since m is arbitrary and the process X is a step process, this implies quasi-left-continuity for $\mathfrak{X}$.

5.21 *Lévy processes.* Suppose that $X - X_0$ is a Lévy process; see Exercise 4.31. Then, $\mathfrak{X}$ is strong Markov. We now show that it is quasi-left-continuous and, hence, a Hunt process.

If X is continuous, there is nothing to prove. Suppose that it has jumps. Fix an integer $m \geq 1$, and consider the successive jump times at which X jumps by an amount whose magnitude is in the interval $[1/m, 1/m-1)$. Those jump times form a Poisson process N; in fact,

$$\mathfrak{X}_m = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X_0 + N, \mathbb{P}^\bullet)$$

is a Markov system and $X_0 + N$ is a simple step process. It follows from the preceding example that $\mathfrak{X}_m$ is quasi-left-continuous. So, every one of its jump times is totally unpredictable.

Since this is true for every $m \geq 1$, and since all those jump times put together exhaust all the jump times of X, we conclude that $\mathfrak{X}$ is quasi-left-continuous.

5.22 *Brown and Poisson.* This is to show the necessity, in Theorem 5.16, of restriction to $\left(\mathcal{G}^o_{t+}\right)$-stopping times. Let $\mathfrak{X}^* = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X^*, \mathbb{P}^\bullet)$ be a Lévy, where $X^* = X_0 + W + N$, with W Wiener, N Poisson, and W and N independent. So, X^* is Hunt. Consider $\mathfrak{X} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X, \mathbb{P}^\bullet)$ where $X = X_0 + W$, which is also a Hunt process. Let T be the time of first jump for the Poisson process N; it is an $\mathcal{F}$-stopping time and it is totally unpredictable (since $\mathfrak{X}^*$ is Hunt). But, for the Brownian motion X, we have $X_{T-} = X_T$. This is possible because T is not a stopping time of $\left(\mathcal{G}^o_{t+}\right)$, the filtration of X itself.

Exiting a holding point

This is to supplement Theorem 4.22 by showing that a strong Markov process exits a holding point only by a jump.

5.23 Proposition. *Suppose that the system $\mathfrak{X}$ has the strong Markov property. Let R be the time of exit from X_0 as in 4.20. Then, for every holding point x in E,*

$$\mathbb{P}^x \{X_{R-} \neq X_R\} = 1.$$

Proof. If x is a trap in E, then $R = \infty$ and $X_{R-} = X_0 = x$ and $X_R = \partial$ almost surely; thus the claim holds trivially. Suppose that x is a holding point but not a trap. Observe that, for every ω in $\{X_0 = x\}$,

$$R(\omega) = r, \quad X_r(\omega) = x \;\Rightarrow\; R(\theta_r \omega) = 0$$

by the definition of R. Thus,

5.24
$$\mathbb{P}^x \{X_R = x,\; R \circ \theta_R = 0\} = \mathbb{P}^x \{X_R = x\}.$$

On the other hand, R is a stopping time of $\left(\mathcal{G}^o_{t+}\right)$ and, therefore, of $(\mathcal{F}_t)$. By the strong Markov property applied at R,

5.25
$$\begin{aligned}\mathbb{P}^x \{X_R = x,\; R \circ \theta_R = 0\} &= \mathbb{E}^x 1_{\{X_R = x\}} \mathbb{P}^{X_R} \{R = 0\} \\ &= \mathbb{E}^x 1_{\{X_R = x\}} \mathbb{P}^x \{R = 0\} = 0\end{aligned}$$

since $\mathbb{P}^x\{R = 0\} = 0$ by the assumption that x is a holding point. It follows from 5.24 and 5.25 that

$$\mathbb{P}^x\{X_R = x\} = 0,$$

which proves the claim since $\mathbb{P}^x\{X_{R-} = x\} = 1$. □

The preceding propositions supplies the rigorous reason for the failure of strong Markov property for Examples 1.28 and 1.29, the delayed uniform motion and Lévy's continuous increasing process. Of course, the proposition has further implications: for instance, if T is a stopping time, on the event that X_T is a holding point, we have $X_R \circ \theta_T \neq X_{R_-} \circ \theta_T = X_T$ almost surely. A somewhat stronger result is next.

No rest for a continuous strong Markov

5.26 Proposition. *Suppose that $\mathfrak{X}$ is strong Markov, and X continuous. Then, almost surely, $t \mapsto X_t$ has no flat segments of finite duration.*

Proof. We are to show that, for almost every ω, there exists no interval $[r,t]$ with $0 \le r < t < \infty$ such that $X_s(\omega) = X_t(\omega)$ for all s in $[r,t]$ and $X_u(\omega) \neq X_t(\omega)$ for some u in (t, ∞). Define

$$Q_t(\omega) = t - inf\,\{r \ge 0 :\ X_s(\omega) = X_t(\omega) \text{ for all } s \text{ in } [r,t]\}\,.$$

If there were such an interval, then there would exist t such that $Q_t(\omega) > \varepsilon$ for some rational number $\varepsilon > 0$ and that $R(\theta_t\omega) < \infty$; note that $R(\theta_t\omega)$ is the length of the interval from t until the exit from $X_t(\omega)$. Hence, with

$$T_\varepsilon = \inf\{t : Q_t > \varepsilon\}\,, \qquad U_\varepsilon = T_\varepsilon + R \circ \theta_{T_\varepsilon},$$

it is enough to show that, for every x in E and every $\varepsilon > 0$,

5.27 $$\mathbb{P}^x\{U_\varepsilon < \infty\} = 0.$$

Fix x and ε such, and drop ε from the notations T_ε and U_ε. The process (Q_t) is adapted to $(\mathcal{G}_t^\circ)$; thus, T is a stopping time of $\left(\mathcal{G}_{t+}^\circ\right)$ and so is U consequently. Observe that, on the event $\{U < \infty\}$ we have, by the definitions of R, T, U,

$$T < \infty,\ R \circ \theta_T > 0,\ X_T = X_U,\ R \circ \theta_U = 0,$$

the last being due to the continuity of X. So,

$$\mathbb{P}^x\{U < \infty\} = \mathbb{P}^x\{U < \infty, R \circ \theta_U = 0\} = \mathbb{E}^x 1_{\{U<\infty\}} \mathbb{P}^{X_U}\{R = 0\}$$

by the strong Markov property at U; and on the event $\{U < \infty\}$,

$$\mathbb{P}^{X_U}\{R = 0\} = \mathbb{P}^{X_T}\{R = 0\} = \mathbb{P}_T^x\{R \circ \theta_T = 0\} = 0$$

since $X_U = X_T$ and $R \circ \theta_T > 0$ on $\{U < \infty\}$. Hence, 5.27 holds. □

The paths are locally bounded

5.28 PROPOSITION. *Suppose that $\mathfrak{X}$ is quasi-left-continuous. Then, for almost every ω and every $t < \zeta(\omega)$, the set $\{X_s(\omega) : 0 \le s \le t\}$ is contained in some compact subset K^ω of E.*

Proof. Since E is locally compact, there is a sequence (K_n) of compact subsets increasing to E and such that K_n is contained in the interior of K_{n+1} for each n. Let T_n be the time of exit from K_n for each n. Then, (T_n) is an increasing sequence of $(\mathcal{G}^\circ_{t+})$-stopping times, and its limit T is again a stopping time. Thus, by the assumed quasi-left-continuity, $\lim X_{T_n} = X_T$ almost surely on $\{T < \infty\}$. By the right-continuity of X and the way the K_n are picked, $X_{T_{n+1}}$ is outside K_n for every n. Hence, the limit X_T is outside E on $\{T < \infty\}$; in other words, $T \ge \zeta$ on $\{T < \infty\}$ and, therefore, on Ω almost surely. Consequently, for almost every ω, if $t < \zeta(\omega)$, then $t < T_n(\omega)$ for some n, in which case $X_s(\omega) \in K_n$ for all $s \le t$.

Strong Markov property

In Definition 5.1, the strong Markov property is stated in its most useful, intuitive form. Several uses of it appeared in the development above. But, how does one tell whether the given system $\mathfrak{X}$ is strong Markov?

For primary processes such as Poisson, Brownian, and Lévy, the strong Markov property was proved directly. For Itô diffusions and jump-diffusions, its proof exploited the dynamics of the motion and the same property for Poisson and Wiener. Next we aim at processes $\mathfrak{X}$ introduced axiomatically; after some preliminaries, we state a condition on (P_t) that ensures both the strong Markov property and the quasi-left-continuity, see Definition 5.36.

5.29 PROPOSITION. *The Markov system $\mathfrak{X}$ is strong Markov if and only if*

5.30 $$\mathbb{E}^x_T\ f \circ X_{T+u} = P_u f \circ X_T, \qquad x \in E,\ u \in \mathbb{R}_+,$$

for every f in $\mathcal{E}_+$ and every stopping time T of $\mathcal{F}$.

Proof. Necessity is obvious. Sufficiency is essentially as in the proof of Theorem 1.2: It is enough to show that 5.30 implies 5.2 for V having the form

$$V_n = f_1 \circ X_{t_1} \cdots f_n \circ X_{t_n}$$

for some $n \ge 1$, times $0 \le t_1 < \ldots < t_n$, and functions $f_1, \ldots, f_n$ in $\mathcal{E}_+$. This is done by induction on n, whose steps are the same as those of the proof of 1.2; basically, replace u_i there with $T + t_i$. We leave out the details. □

5.31 LEMMA. *The system $\mathfrak{X}$ is strong Markov if and only if*

5.32 $$\mathbb{E}^x\ f \circ X_{T+u} = \mathbb{E}^x\ P_u\ f \circ X_T, \qquad x \in E,\ u \in \mathbb{R}_+,$$

for every f in $\mathcal{E}_+$ and every $\mathcal{F}$-stopping time T.

Proof. Applying $\mathbb{E}^x$ to both sides of 5.30 yields 5.32. For the converse, fix f and T, let H be an event in $\mathcal{F}_T$, and consider the stopping time T_H (T on H defined in 5.6). Assuming 5.32, we get

$$\begin{aligned}\mathbb{E}^x 1_H\ f \circ X_{T+u} &= \mathbb{E}^x\ f \circ X_{T_H+u} \\ &= \mathbb{E}^x\ P_u f \circ X_{T_H} = \mathbb{E}^x 1_H\ P_u f \circ X_T.\end{aligned}$$

Since H in $\mathcal{F}_T$ is arbitrary, this is equivalent to 5.30. □

5.33 Lemma. *Let T be an $\mathcal{F}$-stopping time that takes values in a countable subset of $\bar{\mathbb{R}}_+$. Then, the strong Markov property holds at T.*

Proof. For fixed t in $\mathbb{R}_+$,

5.34 $$\mathbb{E}^x 1_{\{T=t\}}\ f \circ X_{t+u} = \mathbb{E}^x 1_{\{T=t\}} P_u f \circ X_t$$

by the Markov property, since $\{T = t\} \in \mathcal{F}_t$. The same holds (and both sides vanish) for $t = +\infty$ as well, via the conventions on X_∞ and $f(\partial)$ and $P_u f(\partial)$. Now, summing both sides of 5.34 over the countably many possible values t for T, we obtain 5.32 via the monotone convergence theorem. □

5.35 Remark. Consider the strong Markov property in the form 5.32. For an arbitrary $\mathcal{F}$-stopping time T, Proposition V.1.20 provides a sequence (T_n) of countably-valued stopping times decreasing to T. By the preceding lemma, 5.32 holds for each T_n. By the right-continuity of X, we have $X_{T_n+u} \to X_{T+u}$ and $X_{T_n} \to X_T$ as $n \to \infty$. Thus, if f and $P_u f$ are continuous and bounded, then 5.32 will hold for T. And, if 5.32 holds for f continuous, then it will hold for all f in $\mathcal{E}_+$ by a monotone class argument, and hence the strong Markov property. For this program to work, we need an assumption that the function $P_u f$ be continuous for f continuous, both regarded as functions on the compact space $\bar{E}$. We take this up next.

Feller processes

Let $\mathcal{C}_0 = C_0(E \mapsto \mathbb{R})$, the set of all continuous functions $f: E \mapsto \mathbb{R}$ with $\lim_{x\to\partial} f(x) = 0$. Elements of $\mathcal{C}_0$ are called continuous functions *vanishing at infinity*. These are functions on E whose automatic extensions (with $f(\partial) = 0$) onto $\bar{E}$ yield continuous functions on $\bar{E}$. Since $\bar{E}$ is compact, every such function is bounded. Every continuous function $\bar{f}: \bar{E} \mapsto \mathbb{R}$ has the form $\bar{f}(x) = f(x) + c$ for some f in $\mathcal{C}_0$ and some constant c, namely, $c = \bar{f}(\partial)$.

5.36 Definition. *The Markov system $\mathfrak{X}$ is called a* Feller system *if*

5.37 $$f \in \mathcal{C}_0 \quad\Rightarrow\quad P_t f \in \mathcal{C}_0 \text{ for every } t \text{ in } \mathbb{R}_+.$$

5.38 Remark. a) Since X is right-continuous, the condition 5.37 implies that

5.39 $$f \in \mathcal{C}_0 \Rightarrow \lim_{t\to 0} P_t f(x) = f(x), \qquad x \in E.$$

For, as $t \to 0$, we have $X_t \to X_0$, and $f \circ X_t \to f \circ X_0$ by the continuity of f; thus

$$P_t f(x) = \mathbb{E}^x\ f \circ X_t \to \mathbb{E}^x\ f \circ X_0 = f(x)$$

by the bounded convergence theorem and the normality of $\mathfrak{X}$.

b) In the absence of $\mathfrak{X}$, a sub-Markov semigroup (P_t) is said to satisfy the *Feller condition* if 5.37 and 5.39 hold. Given such a semigroup, and given E as in 4.1, it is possible to construct a system $\mathfrak{X}$ that satisfies 4.1–4.6 and is a Markov system in the sense of Definition 4.7. The construction is long and tedious. Following the modern sensibilities, we have defined the Markov system $\mathfrak{X}$ axiomatically, rather than treating the semigroup (P_t) (which is rarely explicit except for Wiener and Poisson) as the primary object. The next theorem shows that every Feller process is a Hunt process, that is, it is strong Markov and quasi-left-continuous.

5.40 THEOREM. *If $\mathfrak{X}$ is a Feller system, then it is a Hunt system.*

Proof. Suppose that $\mathfrak{X}$ has the Feller property 5.37, we need to show that, then, $\mathfrak{X}$ is strong Markov and quasi-left-continuous.

a) For the first, we follow the program outlined in Remark 5.35. Let T be an $\mathcal{F}$-stopping time, choose stopping times T_n decreasing to T such that each T_n is countably-valued. By Lemma 5.33,

$$\mathbb{E}^x\ f \circ X_{T_n+u} = \mathbb{E}^x\ P_u\ f \circ X_{T_n}$$

for f in $\mathcal{E}_b$. Now let $f \in \mathcal{C}_0$ and let $n \to \infty$. We obtain 5.32 through the right-continuity of X, the continuity of f and $P_u f$ in $\mathcal{C}_0$ when extended onto $\bar{E}$, and the bounded convergence theorem. Finally, 5.32 extends to f in $\mathcal{E}_+$ by a monotone class argument. Thus, $\mathfrak{X}$ is strong Markov.

b) To show quasi-left-continuity, let T be a stopping time of $\mathcal{F}$, and (T_n) an increasing sequence of such times with limit T; we need to show that

5.41 $$\lim X_{T_n} = X_T \text{ almost surely on } \{T < \infty\}\,.$$

It is enough to show that it is so almost surely on $\{T \le b\}$ for every $b < \infty$; then, letting $b \to \infty$ over the integers yields the desired end. But, on $\{T \le b\}$, we have $T = T \wedge b$ and $T_n = T_n \wedge b$, which are all bounded stopping times. Thus, we may and do assume that T is bounded.

Since X is left-limited in $\bar{E}$, the limits

5.42 $$L = \lim_n X_{T_n}, \qquad L_u = \lim_n\ X_{T_n+u}$$

exist, the latter for every $u > 0$. For $u > 0$, we have $T_n(\omega) + u > T(\omega)$ for all n large enough; thus, $L_u \to X_T$ as $u \to 0$, by the right-continuity of X. Hence, for continuous $\bar{f}$ and $\bar{g}$ on $\bar{E}$,

5.43 $$\mathbb{E}^x\ \bar{f} \circ L \quad \bar{g} \circ X_T = \lim_{u\to 0}\ \lim_{n\to\infty} \mathbb{E}^x\ \bar{f} \circ X_{T_n} \quad \bar{g} \circ X_{T_n+u}.$$

We can write $\bar{g} = c + g$, where $g \in \mathcal{C}_0$ and $c = \bar{g}(\partial)$. Using the already proved strong Markov property at T_n, we see that the right side of 5.43 is equal to

$$\lim_{u \to 0} \lim_{n \to \infty} \mathbb{E}^x \left(\bar{f} \circ X_{T_n}\right) (c + P_u g \circ X_{T_n}) \\ = \lim_{u \to 0} \mathbb{E}^x \left(\bar{f} \circ L\right) (c + P_u g \circ L),$$

where the last equality follows from the bounded continuity of $\bar{f}$ and $P_u g$, the latter being through the assumed Feller property 5.37 applied to g in $\mathcal{C}_0$. Thus, using Remark 5.38a to the effect that $P_u g \to g$ as $u \to 0$, we see that 5.43 becomes,

5.44 $$\mathbb{E}^x \bar{f} \circ L \quad \bar{g} \circ X_T = \mathbb{E}^x \ \bar{f} \circ L \quad \bar{g} \circ L.$$

Since continuous functions of the form $\bar{f} \times \bar{g}$ generate the Borel σ-algebra on $\bar{E} \times \bar{E}$, a monotone class argument applied to 5.44 shows that

$$\mathbb{E}^x \ \bar{h} \circ (L, X_T) = \mathbb{E}^x \ \bar{h} \circ (L, L)$$

for every bounded Borel function $\bar{h}$ on $\bar{E} \times \bar{E}$. Taking $\bar{h}$ to be the indicator of the diagonal of $\bar{E} \times \bar{E}$, and noting the definition of L in 5.42, we obtain the desired result 5.41. □

Markovian bestiary

Poisson processes are the quintessential Markov processes with jumps. Brownian motions are the continuous Markov processes par excellence. They are both Lévy processes.

All Lévy processes are Itô processes; the latter are processes that satisfy stochastic integral equations like 3.1, but with a further term that define a compensated sum of jumps; see 3.95. Itô diffusions, jump-diffusions, and simple processes are special cases of Itô processes.

All Itô processes are Feller processes. The latter are introduced through their transition functions, with conditions on how the transition kernels P_t treat continuous functions. From those conditions follow the real objectives: regularity properties of the sample paths, strong Markov property, quasi-left-continuity, etc.

All Feller processes are Hunt processes. The latter are introduced axiomatically by saying that we have a process and it has the following desirable properties. This is the straightforward approach; it puts the process as the central object, the axioms can be checked directly in practical situations or in the case of Itô processes.

All Hunt processes are "standard;" the latter allow quasi-left-continuity to fail at ζ, at the end of life. Finally, all standard processes are "right processes," the latter form a class of Markov processes that is invariant under certain useful transformations such as killing, time changes, spatial transformations.

These are the objects of the general theory of Markov processes. (See the notes for this chapter for references.)

There is a class of processes that is totally outside of all the previous classes: It consists of Markov processes (in continuous-time) with discrete state spaces, but *without* the sample path regularities such as right-continuity. When the state space is discrete (with the discrete topology), every right-continuous left-limited path is necessarily a step process; too simple, theoretically. On the other hand, on a general state space, it is impossible to build a theory without right-continuity etc. for the paths. But, with a discrete state space, it is possible to create a rich theory that allows sample paths to have discontinuities of the second type. Such processes should be called *Chung processes.*

Exercises and Complements

5.45 *Additive functionals.* Let $\mathfrak{X}$ be a Markov system with living space E. Let $f \in \mathcal{E}_{b+}$ and put

$$A_t = \int_0^t ds\ f \circ X_s, \qquad t \in \mathbb{R}_+.$$

Show that, for every ω,

$$A_{t+u}(\omega) = A_t(\omega) + A_u(\theta_t \omega), \qquad t, u \in \mathbb{R}_+;$$

then, A is said to be *additive.*

5.46 *Continuation.* Let $\mathfrak{X}$ be a Markov system. Let $A = (A_t)$ be an increasing right-continuous process with $A_o = 0$. It is said to be an *additive functional* of X if it is additive and is adapted to $\left(\mathcal{G}^\circ_{t+}\right)$. The preceding exercise gave an example of a continuous additive functional. If X is a Brownian motion, the local time at 0 is another example of a continuous additive functional. If X is a jump-diffusion as in Section 3, then

$$A_t = \sum_{s \le t} f \circ (X_{s-}, X_s)\ 1_{\{X_{s-} \ne X_s\}}, \quad t \in \mathbb{R}_+,$$

is an additive functional of the pure-jump type.

5.47 *Time changes.* Let $\mathfrak{X}$ be a Hunt system with living space E. Let $f : E \mapsto (0, 1)$ be Borel and define

$$C_t = \int_0^t ds\ f \circ X_s, \qquad t \in \mathbb{R}_+.$$

Then, C is a strictly increasing continuous additive functional. Using C as a random clock, let S be its functional inverse (that is, $S_u = \inf\{t \ge 0 : C_t > u\},\ u \in \mathbb{R}_+$). Each S_t is a stopping time of $\left(\mathcal{G}^\circ_{t+}\right)$ and of $(\mathcal{F}_t)$. Define

$$\hat{X}_t = X_{S_t},\ \hat{\theta}_t = \theta_{S_t},\ \hat{\mathcal{F}}_t = \mathcal{F}_{S_t}, \quad t \in \mathbb{R}_+.$$

Show that $\hat{\mathfrak{X}} = (\Omega,\ \mathcal{H},\ \hat{\mathcal{F}},\ \hat{\theta},\ \hat{X},\ \mathbb{P}^\bullet)$ is again a Hunt system with living space E.

5.48 *Increasing continuous processes.* Let $\mathfrak{X}$ be a Hunt system with living space $E = \mathbb{R}_+$. Suppose that $t \mapsto X_t$ is increasing and continuous, and that $\zeta = +\infty$.

a) Show that $t \mapsto X_t$ is strictly increasing.

b) Put $C_t = X_t - X_0$. Show that C is an strictly increasing continuous additive functional of X.

c) Let $\hat{X}$ be the time-changed process. Note that $\hat{X}_t = X_0 + t$, deterministic, except for the initial state. Show that (S_t) is a continuous additive functional of $\hat{X}$. In particular, this means that S_t is determined by $\{\hat{X}_s : s \le t\}$.

d) Conclude that (S_t) and, therefore, (C_t) and (X_t) are deterministic except for the dependence on X_0. Here is the form of X: Let f be a continuous strictly increasing function on $\mathbb{R}_+$ with $f(0) = 0$. If $X_0(\omega) = x$, choose the unique time t_0 such that $f(t_0) = x$; Then $X_t(\omega) = f(t_0 + t)$ for all $t \ge 0$.

5.49 *Step processes.* Let $\mathfrak{X}$ be such that X is a step process; let $(Y_n), (T_n)$, and $k(x)$ be as in 3.69 *et seq.*, with no traps. Let

$$C_t = \int_0^t ds\ k \circ X_s, \qquad t \in \mathbb{R}_+,$$

and consider the process $\hat{X}$ obtained as in 5.47, but from this C.

a) Show that $\hat{X}$ has the form

$$\hat{X}_t = Y_n \qquad on \quad \left\{\hat{T}_n \le t < \hat{T}_{n+1}\right\},$$

where $\hat{T}_0 = 0$, and $\{\hat{T}_{n+1} - \hat{T}_n : n \in \mathbb{N}\}$ is an independency of standard exponential variables that is independent of (Y_n). Thus, $\hat{X}_t = Y_{N_t}$, where N is a standard Poisson process independent of Y.

5.50 *Continuation.* Let $\mathfrak{X}$ be as in Example 5.19 above, where X is a Brownian motion inside $(a,\ b)$ and has sojourns at a and b before jumping into $(a,\ b)$. Define

$$C_t = \int_0^t ds\ 1_{(a,b)} \circ X_s.$$

Note that C remains flat during sojourns of X. Now, C is still a continuous additive functional, but not strictly increasing. Define (S_t) and $\hat{\mathfrak{X}}$ as in the preceding exercise.

a) Show that $\hat{\mathfrak{X}}$ is a Markov system with living space $[a,\ b]$ except that the normality fails (for $\mathbb{P}^a$ and $\mathbb{P}^b$). Of course, the actual state space for $\hat{X}$ is the interval $(a,\ b)$; and $\hat{\mathfrak{X}}$ is a Markov system with living space $(a,\ b)$, since normality does hold for x in $(a,\ b)$ and $x = \partial$.

b) Describe the process $\hat{X}$.

c) Is $\hat{\mathfrak{X}}$ with living space (a,b) a Hunt process?

5.51 *Semimartingale Hunt processes.* Let X be an Itô process; see 3.95. Let $f, C, S, \hat{\mathfrak{X}}$ be as in Exercise 5.47 above. Then, $\hat{\mathfrak{X}}$ is a Hunt system as mentioned in 5.47. Moreover, every component $\hat{X}^i$ of $\hat{X}$ is a semimartingale.

This has a converse. Every Hunt process $\hat{X}$ whose components are semimartingales has the structure described. Somewhat more explicitly, let $\hat{X}$ be a Hunt process with state space $E = \mathbb{R}^d$. Then, there are deterministic measurable functions a, b, j, f and (on an enlargement of the original probability space) a Wiener process W and a Poisson random measure M such that $\hat{X}_t = X_{S_t}$ where X is an Itô process as described in 3.95 and S is defined from X and f as in 5.47. See the notes for this chapter.

6 Potentials and Excessive Functions

This section is independent of Section 5, and its dependence on the earlier sections is slight. Throughout, $\mathfrak{X} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, X, \mathbb{P}^\bullet)$ is a Markov system with living space E and transition semigroup (P_t); see 4.1–4.6 and Definition 4.7 for these and the attendant conventions.

For f in $\mathcal{E}_b$ and $p > 0$, by the arguments of Example 4.18,

6.1 $$U_p f(x) = \mathbb{E}^x \int_0^\infty dt\, e^{-pt} \quad f \circ X_t, \qquad x \in \mathbb{E},$$

defines a function $U_p f$ in $\mathcal{E}_b$. The same makes sense for f in $\mathcal{E}_+$ and $p \geq 0$, and the result is a function $U_p f$ in $\mathcal{E}_+$. In both cases,

6.2 $$U_p f(x) = \int_E \quad U_p(x, dy)\ f(y), \qquad x \in E,$$

where

6.3 $$U_p(x, A) = \mathbb{E}^x \int_0^\infty dt\, e^{-pt}\ 1_A \circ X_t = \int_0^\infty dt\, e^{-pt}\ P_t(x, A).$$

If $p > 0$, then U_p is a bounded kernel: $U_p(x, E) \leq {}^1/_p$. When $p = 0$, writing the integral over t as a sum of integrals over $[n,\ n+1)$ shows that U_0 is Σ-bounded, but generally not σ-finite.

The function $U_p f$ is called the *p-potential* of f, and U_p the *p-potential kernel* or *p-potential operator* depending on the role it plays. The family $(U_p)_{p>0}$ of operators $U_p : \mathcal{E}_b \mapsto \mathcal{E}_b$ is called the *resolvent* of the semigroup (P_t) or of the Markov process X.

6.4 Theorem. *a) The resolvent equation*

6.5 $$U_p - U_q + (p - q)\, U_p\ U_q = 0$$

holds for $p, q > 0$; in particular, $U_p\ U_q = U_q\ U_p$.

b) For each $p > 0$, the kernel pU_p is sub-Markov; and

$$\lim_{p \to \infty} pU_p\ f(x) = f(x)$$

for every f in $\mathcal{E}_b$ that is continuous at the point x of E.

Proof. Follows from the same arguments as in 2.36–2.39 and Theorem 2.40. □

Potentials and supermartingales

6.6 THEOREM. *Let $p > 0$ and $f \in \mathcal{E}_{b+}$. Then, for each x in E,*

$$M_t = \int_0^t ds\ e^{-ps}\ f \circ X_s + e^{-pt}\ U_p\ f \circ X_t, \qquad t \in \mathbb{R}_+,$$

is a uniformly integrable $\mathcal{F}$-martingale over $(\Omega, \mathcal{H}, \mathbb{P}^x)$.

Proof. Define

6.7 $$A_t = \int_0^t ds \quad e^{-ps}\ f \circ X_s, \qquad t \in \mathbb{R}_+.$$

The process (A_t) is increasing, and the limit A_∞ is a bounded positive variable in $\mathcal{G}_\infty$ since $p > 0$ and f is bounded. Note that

$$A_\infty = A_t + \int_t^\infty ds\ e^{-ps}\ f \circ X_s = A_t + e^{-pt}\ A_\infty \circ \theta_t.$$

Thus, by the Markov property at t,

$$\mathbb{E}_t^x\ A_\infty = A_t + e^{-pt}\ \mathbb{E}^{X_t}\ A_\infty = A_t + e^{-pt}\ U_p f \circ X_t = M_t,$$

since $\mathbb{E}^y A_\infty = U_p f(y)$ by definition. Via Theorem V.5.13, this shows that M is a uniformly integrable martingale, with respect to $\mathcal{F}$, under every $\mathbb{P}^x$. □

6.8 COROLLARY. *Under each $\mathbb{P}^x$,*

$$V_t = e^{-pt}\ U_p f \circ X_t, \qquad t \in \mathbb{R}_+,$$

is a positive supermartingale with $\lim_{t\to\infty} V_t = 0$.

Proof. We have $V = M - A$ with the definitions in 6.6 and 6.7, and the process A is increasing. Thus, V is a supermartingale. And, $\lim_{t\to\infty} V_t = 0$, because, by the martingale convergence theorem,

$$\lim_{t\to\infty} M_t = \lim_t \mathbb{E}_t^x\ A_\infty = A_\infty = \lim_{t\to\infty} A_t.$$ □

In martingale terminology, the process (V_t) is a potential; see V.4.53. The decomposition

$$V = M - A$$

is an instance of *Doob-Meyer decomposition* for supermartingales, which is the continuous-time version of Doob's decomposition given in Theorem V.3.2.

Going back to Theorem 6.6, the uniform integrability of M implies that M is a Doob martingale on $[0, \infty]$, and thus

$$\mathbb{E}^x\ M_T = \mathbb{E}^x M_0 = U_p f(x)$$

for every stopping time T of $\mathcal{F}$. This proves the following corollary to 6.6.

6.9 THEOREM. *Let $p > 0$ and $f \in \mathcal{E}_{b+}$. Then, for every x in E,*

$$U_p f(x) = \mathbb{E}^x \int_0^T ds\ e^{-ps}\ f \circ X_s\ +\ \mathbb{E}^x\ e^{-pT}\ U_p f \circ X_T$$

for every $\mathcal{F}$-stopping time T. In particular, if $f \circ X_s = 0$ on $\{s < T\}$, then

$$\mathbb{E}^x\ e^{-pT}\ U_p\ f \circ X_T = U_p\ f(x), \quad x \in E.$$

The particular case is useful in computing the distributions of T and X_T by choosing f appropriately. The theorem is the potential counterpart of Dynkin's formula using generators; see Theorem 2.27 for Dynkin's formula for Itô diffusions.

Excessive functions

6.10 DEFINITION. *Let $p \in \mathbb{R}_+$. A function f in $\mathcal{E}_+$ is said to be p-excessive if*

a) $f \geq e^{-pt}\ P_t f$ for every t in $\mathbb{R}_+$, and

b) $\lim_{t \to 0} e^{-pt}\ P_t f(x) = f(x)$ for every x in E.

The condition (a) is called the *p-supermedian property* for f; other terms in use are *p-super-mean-valued*, *p-superaveraging*. It implies, via the semigroup property $P_t P_u = P_{t+u}$, that the mapping $t \mapsto e^{-pt} P_t f(x)$ is decreasing. Hence, the limit in (b) is an increasing limit, and the condition (b) can be written as

6.11 $$\sup_t\ e^{-pt}\ P_t\ f(x) = f(x).$$

If the conditions (a) and (b) hold for Borel f, without requiring that f be positive, then f is said to be *p-superharmonic*. In all this, when $p = 0$, it is dropped both from notation and terminology. The following is the connection to supermartingales.

6.12 PROPOSITION. *Let $p \geq 0$. Let f be p-supermedian. Then, for each x in E with $f(x) < \infty$, the process*

$$Y_t = e^{-pt}\ f \circ X_t, \qquad t \in \mathbb{R}_+,$$

is an $\mathcal{F}$-supermartingale over $(\Omega, \mathcal{H}, \mathbb{P}^x)$.

Proof. Obviously, Y is adapted to $\mathcal{F}$. Fix x such that $f(x) < \infty$. Then, since f is p-supermedian,

$$f(x) \geq e^{-pt}\ P_t f(x) = \mathbb{E}^x\ e^{-pt}\ f \circ X_t = \mathbb{E}^x\ Y_t,$$

showing that Y_t is integrable under $\mathbb{P}^x$. And, by the Markov property of X,

$$\begin{aligned} \mathbb{E}_t^x\ Y_{t+u} &= e^{-p(t+u)}\ \mathbb{E}_t^x\ f \circ X_{t+u} \\ &= e^{-pt}\ e^{-pu}\ P_u\ f \circ X_t \leq e^{-pt}\ f \circ X_t = Y_t, \end{aligned}$$

where the inequality is via the p-supermedian property of f. □

In the preceding proposition, if f is p-excessive, it can be shown that Y is right-continuous. Thus, p-excessive functions are, roughly speaking, continuous over the paths of X.

Potentials are excessive

6.13 PROPOSITION. *Let $p \geq 0$ and $f \in \mathcal{E}_+$. Then, $U_p f$ is p-excessive.*

Proof. Clearly, $U_p f \in \mathcal{E}_+$. Also, by Fubini's theorem,

$$e^{-pt}\ P_t\ U_p\ f = e^{-pt} \int_0^\infty du\ e^{-pu}\ P_t\ P_u\ f = \int_t^\infty ds\ e^{-ps}\ P_s f.$$

The last integral is dominated by $U_p f$ and increases to $U_p f$ as t decreases to 0. Hence, $U_p f$ is p-excessive. □

The following is one-half of a theorem characterizing excessive functions in terms of the resolvent. But it is sufficient for our purposes.

6.14 PROPOSITION. *Let $p \geq 0$ and let f be p-supermedian. Then $q \mapsto qU_{p+q} f$ is increasing and dominated by f. Its limit is f as $q \to \infty$ if f is p-excessive.*

Proof. Fix $p \geq 0$. For $q > 0$,

$$q\ U_{p+q}\ f = \int_0^\infty dt\ q\ e^{-qt}\ e^{-pt}\ P_t f = \int_0^\infty du\ e^{-u} e^{-pu/q}\ P_{u/q}\ f.$$

As q increases, u/q decreases and the integrand increases by the p-supermedian property of f. By the same property, the last integrand is dominated by $e^{-u} f$, and hence, the integral is dominated by f. Finally, if f is p-excessive, the integrand increases to $e^{-u} f$ as $q \to \infty$, and the monotone convergence theorem implies that the integral becomes f in the limit. □

Approximation by bounded potentials

Let f be p-excessive. Then, for each n, the function $nU_{p+n}f = U_{p+n}(nf)$ is a potential, and as $n \to \infty$ the limit is f. So, every p-excessive function is the limit of an increasing sequence of potentials. The following sharpens the result when $p > 0$.

6.15 THEOREM. *Let $p > 0$. Let f be p-excessive. Then, there exists a sequence (g_n) in $\mathcal{E}_{b+}$ such that the sequence $(U_p g_n)$ increases to f.*

Proof. For each integer $n \geq 1$, put $f_n = f \wedge n$; each f_n is bounded and p-supermedian (since f is such and the constant n is such a function). By the resolvent equation 6.5,

$$U_{p+q}\ f_n = U_p\ f_n - q\ U_p\ U_{p+q}\ f_n;$$

the right side is well-defined as the difference of two bounded functions since f_n is bounded and $p > 0$. Thus, with $g_n = n(f_n - nU_{p+n}f_n)$, we have

6.16 $$nU_{p+n}\ f_n = U_p\ g_n.$$

Since f_n is p-supermedian, Proposition 6.14 yields $f_n \geq nU_{p+n}f_n$, and hence $g_n \in \mathcal{E}_{b+}$ for every n. There remains to show that the left side of 6.16 increases to f as $n \to \infty$. To that end, we note that (f_n) is increasing to f, and that $qU_{p+q}f_n$ is increasing in q, since f_n is p-supermedian (see Proposition 6.14). Thus, the left side of 6.16 is increasing in n, and

$$\lim_n\ nU_{p+n}\ f_n = \lim_q\ \lim_n\ qU_{p+q}\ f_n = \lim_q\ q\ U_{p+q}\ f = f,$$

the last equality being via Proposition 6.14 applied to the p-excessive function f. □

Supermedian property at stopping times

This is essentially Doob's stopping theorem for the supermartingale Y of Proposition 6.12. See Exercises 6.23–6.26 for its interpretation in optimal stopping games.

6.17 THEOREM. *Let $p \geq 0$. Let f be p-excessive. Then, for every $\mathcal{F}$-stopping time T,*

6.18 $$f(x)\ \geq\ \mathbb{E}^x\ e^{-pT}\ f \circ X_T, \qquad x \in E.$$

Proof. Suppose that $p > 0$. Let (g_n) be as in Theorem 6.15, so that $U_p g_n \nearrow f$. By Theorem 6.9,

$$U_p\ g_n(x) \geq \mathbb{E}^x\ e^{-pT}\ U_p g_n \circ X_T.$$

Letting $n \to \infty$ we obtain 6.18 when $p > 0$.

If $p = 0$ and f is excessive, then f is p-excessive for every $p > 0$, and thus 6.18 holds for every $p > 0$. Let p decrease to 0 strictly. By the monotone convergence theorem,

$$f(x) \geq \lim_{p \to 0} \ \mathbb{E}^x \ e^{-pT} \ f \circ X_T = \mathbb{E}^x \ 1_{\{T<\infty\}} f \circ X_T = \mathbb{E}^x \ f \circ X_T,$$

since $X_T = \partial$ on $\{T = \infty\}$ and $f(\partial) = 0$.

Exercises

6.19 *Poisson process.* Let $X = X_0 + N$ where N is a Poisson process with rate c; we take $E = \mathbb{R}_+$. Show that

$$Uf(x) = \mathbb{E}^x \int_0^\infty dt \ f \circ X_t = \frac{1}{c} \sum_{j=0}^{\infty} f(x+j), \qquad x \in E,$$

for every f positive Borel, more generally, for $p \geq 0$, show that

$$U_p f(x) = \frac{1}{c+p} \sum_{j=0}^{\infty} \left(\frac{c}{c+p} \right)^j f\,(x+j)\,.$$

6.20 *Stable processes.* Suppose that $X = X_0 + S$, where S is an increasing stable process with index a in (0,1). Suppose that its Lévy measure is given as $\lambda(dx) = (c/x^{a+1})dx$, with $c = a/\Gamma(1-a)$; see Example VII.7.6b. Show that, with $E = \mathbb{R}_+$,

$$Uf(x) = \frac{1}{\Gamma(a)} \int_x^\infty dy \ (y-x)^{a-1} \ f(y), \qquad x \in E.$$

6.21 *Brownian motion.* Suppose that $X = X_0 + W$, the standard Brownian motion in $\mathbb{R}^d$.

a) For $d = 1$ or $d = 2$, show that

$$Uf\ (x) = +\infty, \qquad x \in \mathbb{R}^d,$$

for every f positive Borel on $\mathbb{R}^d$, except for $f = 0$ in which case $Uf = 0$.

b) Show, if $d \geq 3$, that

$$Uf(x) = \frac{\Gamma\left(\frac{d}{2}-1\right)}{2\ \pi^{d/2}} \int_{\mathbb{R}^d} dy\, |y-x|^{2-d} \ f(y)$$

for all x in $\mathbb{R}^d$ and positive Borel f on R^d. Thus, except for multiplication by a constant, Uf is the Newtonian potential of f in classical potential theory.

6.22 *Excessive functions.* For the Markov system $\mathfrak{X}$, prove the following.

a) If $f = c$ for some constant $c \geq 0$, then f is p-excessive for every $p \geq 0$.

b) If f is p-excessive, then it is q-excessive for every $q \geq p$.

c) If f is p-excessive and $c \geq 0$ is a constant, then cf is p-excessive. If f and g are p-excessive, then so is $f + g$.

d) If (f_n) is an increasing sequence of p-excessive functions with limit f, then f is p-excessive. Hint: $e^{-pt}\ P_t\ f_n$ is increasing in n, and increasing with decreasing t.

e) If f and g are p-supermedian, then so is $f \wedge g$.

6.23 *Brownian motion on* $\mathbb{R}$. Let X be a standard Brownian motion on $\mathbb{R}$. Let f be excessive ($p = 0$). Show that $f = c$ for some constant $c \geq 0$. Hint: Use 6.18 with $T = T_y$, the hitting time of the point y.

6.24 *Continuation.* Let $\hat{X}$ be a standard Brownian motion on $\mathbb{R}$, and let $X_t = \hat{X}_{\tau \wedge t}$, where τ is the time of exit for $\hat{X}$ from the fixed interval (a, b). Thus, X lives in $E = [a, b]$, and the boundary points a and b are traps. Show that every excessive function for X is a concave function on $[a, b]$. Hint: Recall the formula for $\mathbb{E}^x\ f \circ X_T$ for $y < x < z$ and T the time of exit from the interval $(y, z) \subset [a, b]$. Fix y and z, take $x = \alpha y + (1 - \alpha)z$ for $0 \leq \alpha \leq 1$.

6.25 *Optimal stopping.* We are to receive a one-time reward of $f \circ X_T$ dollars if we choose time T to ask for the reward. We want to choose a stopping time T_o that maximizes

$$\mathbb{E}^x\ e^{-rT}\ f \circ X_T, \qquad T \text{ is an } \mathcal{F}\text{-stopping time},$$

if possible, or, if this proves impossible, come close to the value

$$v(x) = \sup_T \mathbb{E}^x\ e^{-rT}\ f \circ X_T,$$

where the supremum is over all $\mathcal{F}$-stopping times. Here, f is a positive Borel function on E, called the payoff function. We interpret r as the interest rate, and v is called the value of the game.

a) If f is r-excessive, then $v = f$ and $T_0 = 0$ is an optimal stopping time.

b) In general, v is the minimal r-excessive function that dominates f.

6.26 *Continuation.* Suppose that X is the standard Brownian motion on $\mathbb{R}$. Let f be a bounded positive function on $\mathbb{R}$, and take $r = 0$. Show that $v = c$, no computations needed, where

$$c = \sup_{y \in \mathbb{R}}\ f(y).$$

If the supremum is attained, that is, there exists x^* in $\mathbb{R}$ such that $f(x^*) = c$, then the time T_0 of hitting x^* is an optimal stopping time. If $f(x) = 1 - e^{-x}$ for $x > 0$ and is 0 otherwise, then $v(x) = c = 1$ for all x in $\mathbb{R}$; but there is no optimal stopping time; recall that $X_\infty = \partial$ and $f(\partial) = 0$ – the dead pay nothing. As this example indicates, there might be no optimal stopping time. But, for every $\varepsilon > 0$ there is a stopping time T_ε such that

$$\mathbb{E}^x\ f \circ X_{T_\varepsilon} \geq v(x) - \varepsilon.$$

7 Appendix: Stochastic Integration

This is a quick introduction to stochastic calculus. It is driven by the needs of Section 2 on Itô diffusions. We limit ourselves mostly to continuous processes and omit almost all proofs.

Throughout, $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space, and $\mathcal{F} = (\mathcal{F}_t)$ is an augmented right-continuous filtration. All processes are adapted to this $\mathcal{F}$, without further mention. Also, all processes have the state space $(\mathbb{R},\ \mathcal{B}_\mathbb{R})$. We introduce the term *Stieltjes process* as a short substitute for a process whose almost every path is right-continuous, is left-limited, and has bounded total variation over every bounded interval. Then, to repeat Definition V.5.18, a process X is a semimartingale if it can be written as the sum of a local martingale and a Stieltjes process.

Stochastic integrals

For our current purposes, we call $\sigma = (t_i)$ a *subdivision* of $\mathbb{R}_+$ if $0 = t_0 < t_1 < \cdots$ and $\lim\ t_n = +\infty$. A subdivision of $[0,t]$ is a finite sequence $\sigma = (t_0,\ t_1, \ldots, t_n)$ with $0 = t_0 < t_1 < \cdots < t_n = t$. In both cases, $\|\sigma\| = \sup_i\ (t_{i+1} - t_i)$ is called the mesh of σ.

Let F be a left-continuous process, and X continuous. For every subdivision $\sigma = (t_i)$ of $\mathbb{R}_+$, we define a new process Y^σ by putting $Y_0^\sigma = 0$ and

$$7.1\quad Y_t^\sigma = \sum_{i=1}^{j} F_{t_{i-1}} \cdot \left(X_{t_i} - X_{t_{i-1}}\right) + F_{t_j} \cdot \left(X_t - X_{t_j}\right) \qquad \text{if } t_j < t \le t_{j+1}.$$

Then, Y^σ is a continuous process. Note the resemblance to V.3.4, the integral in discrete time. We omit the proof of the following fundamental result.

7.2 Theorem. *Let F be a left-continuous process, and X a continuous semimartingale. Then, there exists a unique process Y such that*

$$\lim_{\|\sigma\| \to 0} \mathbb{P}\left\{ \sup_{0 \le t \le u} |Y_t^\sigma - Y_t| > \varepsilon \right\} = 0$$

for every $\varepsilon > 0$ and $u < \infty$. The process Y is a continuous semimartingale. □

7.3 Definition. *The process Y of the preceding theorem is called the* stochastic integral *of F with respect to X, and the notations*

$$\int F\, dX \quad \text{and} \quad \int_0^t F_s\, dX_s$$

are used, respectively, for the process Y and the random variable Y_t.

7.4 REMARK. a) The theorem can be re-phrased: as $\|\sigma\| \to 0$, $Y_t^\sigma \to Y_t$ in probability, uniformly in t over compacts. The uniqueness of Y is up to indistinguishability.

b) If X is a Stieltjes process, then the stochastic integral coincides with the path-by-path ordinary integral, that is, for almost every ω, the number $Y_t(\omega)$ is the Riemann-Stieltjes integral of the function $s \mapsto F_s(\omega)$ with respect to the bounded variation function $s \mapsto X_s(\omega)$ over the interval $[0,t]$. Of course, then Y is a continuous Stieltjes process.

c) If X is not Stieltjes, if $X = W$ Wiener for instance, then $Y_t(\omega)$ is not the limit of $Y_t^\sigma(\omega)$ with ω held fixed. In fact, in most cases, $\lim_{\|\sigma\|\to 0} Y_t{}^\sigma(\omega)$ will not exist.

7.5 EXAMPLE. *Wiener driven integrals.* Suppose that F is left-continuous and bounded, and $X = W$, a Wiener process. Then, Y is an L^2-martingale and

7.6
$$\mathbb{E}|\int_0^t F_s \; dW_s|^2 = \mathbb{E}\int_0^t |F_s|^2 \; ds.$$

Here is the explanation. Given a subdivision $\sigma = (t_i)$, define the left-continuous step process F^σ by letting $F_t^\sigma = F_{t_{i-1}}$ for $t_{i-1} < t \le t_i$, and $F_0^\sigma = F_0$. Note that, in fact, Y^σ is the integral of F^σ with respect to X by every reasonable definition of integration. It is evident from 7.1 that Y^σ is now a martingale with

7.7
$$E\,|Y_t^\sigma|^2 = \mathbb{E}\int_0^t |F_s^\sigma|^2\, ds;$$

this is an easy computation recalling that the increments $W_{t_i} - W_{t_{i-1}}$ are independent with mean 0 and variances $t_i - t_{i-1}$. In fact, since F is bounded, $Y_t^\sigma \to Y_t$ in the sense of L^2-convergence as $\|\sigma\| \to 0$. And, $F^\sigma \to F$ by the left-continuity of F. Thus, letting $\|\sigma\| \to 0$ on both sides of 7.7 we obtain 7.6.

Arithmetic of integration

Stochastic integrals are the same as ordinary integrals in linearity etc. The next proposition shows them; proofs are immediate from 7.1–7.3.

7.8 THEOREM. *Let F and G be left-continuous processes, X and Y continuous semimartingales, and a and b constants. Then,*

$$\int (aF + b\,G)\; dX = a\int F\; dX + b\int G\; dX,$$
$$\int F\; d(aX + bY) = a\int F\; dX + b\int F\; dY,$$
$$Y = \int F\; dX \Rightarrow \int G\; dY = \int (F\cdot G)\; dX.$$

7.9 REMARK. Let X be a continuous semimartingale. Then,

7.10
$$X = L + V,$$

where L is a continuous local martingale and V is a continuous Stieltjes process. For F left-continuous, then

7.11
$$\int F \, dX = \int F \, dL + \int F \, dV,$$

and, on the right side, the first term is a continuous local martingale, and the second is a continuous Stieltjes. In particular, if L is a martingale and F is bounded, then the first term is a martingale; Example 7.5 is a special case.

Cross variation, quadratic variation

Given processes X and Y, and a subdivision $\sigma = (t_i)$ of $\mathbb{R}_+$, let

7.12
$$C_t^\sigma = \sum_{t_i < t} \left(X_{t_{i+1}} - X_{t_i} \right) \left(Y_{t_{i+1}} - Y_{t_i} \right), \qquad t \in \mathbb{R}_+.$$

When $X = Y = W$, Wiener, we have seen in Theorem VIII.7.2 that $C_t^\sigma \to t$ in probability as $\|\sigma\| \to 0$. The following is the general case.

7.13 THEOREM. *Let X and Y be continuous semimartingales. Then, there is a continuous Stieltjes process C such that $C_t^\sigma \to C_t$ in probability for every t in the limit as $\|\sigma\| \to 0$.* □

The process C of the preceding theorem is called the *cross variation* of X and Y, and the notation $\langle X, Y \rangle$ is employed for it, that is,

7.14
$$\langle X, Y \rangle_t = C_t, \qquad t \in \mathbb{R}_+.$$

In particular, $\langle X, X \rangle$ is called the *quadratic variation* for the continuous semimartingale X; it is an increasing process in view of 7.12. The approximation 7.12 shows as well that

7.15
$$\langle X + Y, \ X + Y \rangle = \langle X, X \rangle + 2\langle X, Y \rangle + \langle Y, Y \rangle.$$

Solving this for $\langle X, Y \rangle$, since all other terms are increasing processes, we see that $\langle X, Y \rangle$ is indeed the difference of two increasing processes (as a Stieltjes process must be).

7.16 PROPOSITION. *Let X and Y be continuous semimartingales; if X or Y is Stieltjes, then $\langle X, Y \rangle = 0$. In particular, if $X = L + V$ as in 7.10, then $\langle X, X \rangle = \langle L, L \rangle$.*

Proof. Suppose Y is Stieltjes. From 7.12, we have

$$|C_t^\sigma| \le \sup_{t_j < t} \left| X_{t_{j+1}} - X_{t_j} \right| \sum_{t_i < t} \left| Y_{t_{i+1}} - Y_{t_i} \right|.$$

As $\|\sigma\| \to 0$, the supremum goes to 0 by the continuity of X, and the sum goes to the total variation of Y over $[0,t]$. The latter is finite since Y is assumed to be Stieltjes. Thus, $C_t^\sigma \to 0$, that is, $\langle X, Y\rangle = 0$. The particular statement follows from 7.15 for $\langle L+V,\ L+V\rangle$, because $\langle L, V\rangle = \langle V, V\rangle = 0$ since V is continuous Stieltjes.

Stochastic differentials

In analogy with ordinary calculus, we write

7.17 $$dY = F\ dX \ \Leftrightarrow\ Y_t = Y_0 + \int_0^t F_s\ dX_s, \quad t \in \mathbb{R}_+.$$

Similarly, in view of 7.12–7.14, we introduce the notation

7.18 $$dX\ dY = d\langle X, Y\rangle.$$

In particular, $dX\ dX = (dX)^2$ becomes the notation for the differential of the increasing continuous process $\langle X, X\rangle$. Next are the rules of stochastic differential calculus.

7.19 THEOREM. *Let F and G be left-continuous processes, X and Y continuous semimartingales, and a and b constants. Then,*

a) $d(aX + bY) = a\ dX + b\ dY$,

b) $F \cdot (dX + dY) = F\ dX + b\ dY$,

c) $(aF + b\ G)\ dX = a\ F\ dX + b\ G\ dX$,

d) $F \cdot (G\ dX) = (F \cdot G)\ dX$,

e) $(F\ dX)(G\ dY) = (F \cdot G)\ dX\ dY$, *and*

f) if X or Y is Stieltjes, then $dX\ dY = 0$.

Proof. (a) is direct from the definitions; (b), (c), (d) are the differential versions of the properties listed in Theorem 7.8; (e) follows by a simple computation from 7.12 upon replacing X there with $\int F\ dX$, and Y with $\int G\ dY$; finally, (f) is a re-statement of Proposition 7.16.

Itô's formula

This is the chain rule of differentiation for stochastic calculus. Recall that $C^2(\mathbb{R}^d \mapsto \mathbb{R})$ is the class of function $f : \mathbb{R}^d \mapsto \mathbb{R}$ that are twice continuously differentiable, and that we write $\partial_i f(x)$ for $\frac{\partial}{\partial x_i} f(x)$, and $\partial_{ij} f(x)$ for $\frac{\partial^2}{\partial x_i \partial x_j} f(x)$. When $d = 1$, we write f' for the first derivative, and f'' for the second. The next is Itô's formula for continuous semimartingales.

7.20 THEOREM. *Let $X^1, \ldots, X^n$ be continuous semimartingales, and put $X = (X^1, \ldots, X^n)$. For f in $C^2(\mathbb{R}^n \mapsto \mathbb{R})$, then, $f \circ X$ is a semimartingale, and*

$$d(f \circ X) = \sum_{i=1}^{n} (\partial_i f \circ X)\ dX^i\ + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n}\ (\partial_{ij} f \circ X)\ \ dX^i\, dX^j.$$

7.21 REMARK. a) For $n = 1$, Itô's formula becomes

$$d(f \circ X) = (f' \circ X)\ \ dX\ + \frac{1}{2} (f'' \circ X)\ \ (dX)^2.$$

b) If $X^1, \ldots, X^n$ are continuous Stieltjes processes, then $dX^i\ dX^j = 0$ for all i and j, and Itô's formula becomes the chain rule of differential calculus:

$$d\,(f \circ X) = \sum_{i=1}^{n} (\partial_i f \circ X)\ \ dX^i.$$

c) We re-state the conclusion of the theorem above in the formal notation of stochastic integrals:

$$f \circ X_t = f \circ X_0 + \sum_{i=1}^{n} \int_0^t (\partial_i f \circ X_s)\ \ dX_s^i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int_0^t (\partial_{ij} f \circ X_s)\, d\langle X^i, X^j \rangle_s.$$

d) Taking $n = 2$ and $f(x, y) = xy$ in the preceding formula, we obtain the following formula for integration by parts for continuous semimartingales X and Y:

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s\ dY_s + \int_0^t Y_s\ dX_s + \langle X, Y \rangle_t.$$

Wiener driven integrals

7.22 LEMMA. *Let X and Y be independent Wiener processes. Then, $dX\ dY = 0$.*

Proof. We have $\langle X, X \rangle_t = \langle Y, Y \rangle_t = t$ by Theorem VIII.7.2. By the same theorem, since $X + Y = \sqrt{2}\, W$ for some Wiener W, we have $\langle X + Y, X + Y \rangle_t = 2t$. The claim now follows from 7.15. □

Adding the preceding lemma to Itô's formula proves the following.

7.23 THEOREM. *Let $W^1, \ldots, W^n$ be independent Wiener processes and put $W = (W^1, \ldots, W^n)$. For f in $C^2(\mathbb{R}^n \mapsto \mathbb{R})$, then,*

$$d\,(f \circ W_t) = \sum_{i=1}^{n} (\partial_i f \circ W_t)\, dW_t^i + \frac{1}{2} \sum_{i=1}^{n} (\partial_{ii} f \circ W_t)\ \ dt.$$

Characterizations for Wiener processes

Recall the martingale characterization for Wiener processes; see Theorem V.2.19 and Proposition V.6.21: The processes W and $Y = \left(W_t^2 - t\right)$ are continuous martingales with $W_0 = Y_0 = 0$ if and only if W is a Wiener process. Itô's formula in 7.20 with $n = 1$ and $f(x) = x^2$ identifies the process Y,

$$W_t^2 - t = 2\int_0^t W_s \, dW_s,$$

and shows, furthermore, that $\left(W_t^2 - t\right)$ is a martingale if and only if $\langle W, W\rangle_t = t$. This last property characterizes Wiener processes among all continuous local martingales:

7.24 THEOREM. *Let X be a continuous local martingale with $X_0 = 0$. Then, X is a Wiener process if and only if $\langle X, X\rangle_t = t$ for all $t \geq 0$.*

Proof. Necessity is by Theorem VIII.7.2. We show the sufficiency next. Suppose that $\langle X, X\rangle_t = t$ for all t. Then, by Itô's formula with $n = 1$ and f in $C^2(\mathbb{R} \longmapsto \mathbb{R})$,

$$f \circ X_t = f(0) + \int_0^t (f' \circ X_s)\, dX_s + \frac{1}{2}\int_0^t (f'' \circ X_s)\, ds.$$

Assuming further that f, f', f'' are bounded, the stochastic integral term on the right side defines a martingale. This is the content of Lemma V.6.22, and the proof of Proposition V.6.21 applies to show that X is a Wiener process. □

The next theorem is the n-dimensional version of the preceding. The necessity part of its proof is by Lemma 7.22; we omit the proof of sufficiency (it is similar to that of V.6.21).

7.25 THEOREM. *Let $X^1, \ldots, X^n$ be continuous local martingales with $X_0^i = 0$ for every i. Then, $X^1, \ldots, X^n$ are independent Wiener processes if and only if*

$$\langle X^i, X^j\rangle_t = I\,(i,j)\; t, \qquad t \in \mathbb{R}_+,$$

where I is the identity matrix in n-dimensions.

Itô's formula and the characterization above in terms of cross variations form a summary of stochastic integrals driven by Wiener processes.

Local martingales as stochastic integrals

If $dX = F\, dW$, then $(dX)^2 = F^2\,(dW)^2 = F^2\, dt$. The following theorem provides a converse as well.

7.26 THEOREM. *Let X be a continuous local martingale. Suppose that $(dX_t)^2 = (F_t)^2\, dt$ for some left-continuous process F. Then, there is a*

Wiener process W (possibly on an enlargement of the original probability space) such that

$$X_t = X_0 + \int_0^t F_s \, dW_s$$

Proof. For the case $F > 0$. Then, since $1/F$ is left-continuous,

$$dW = \frac{1}{F} \, dX, \qquad W_0 = 0,$$

defines a continuous local martingale W. Since $(dW)^2 = (1/F)^2 \, (dX)^2 = dt$ by the assumption on $(dX)^2$, we see from the characterization theorem 7.24 that W is Wiener. Obviously, $dX = F \, dW$ as claimed. □

Local martingales are time changed Wieners

7.27 THEOREM. *Let X be a continuous local martingale. Let $C = \langle X, X \rangle$. Then, there is a Wiener process W such that*

$$X_t = X_0 + W_{C_t}, \qquad t \in \mathbb{R}_+.$$

Proof is omitted, but its essentials can be seen in Figure 17 below. If time is reckoned with the random clock $C = \langle X, X \rangle$, then $X - X_0$ appears as a Wiener process.

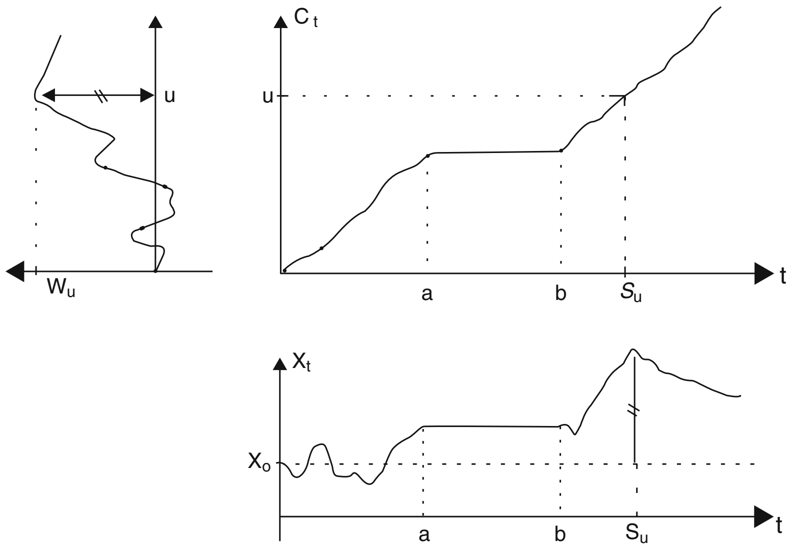


Figure 17: Using $C = \langle X, X \rangle$ as a random clock converts the local martingale X into a Wiener process. Conversely, X is obtained from the Wiener process by reversing the procedure.

For the outcome ω pictured in Figure 17, the path $t \mapsto C_t(\omega)$ remains flat over the interval $[a(\omega),\ b(\omega)]$, and then, $t \mapsto X_t(\omega) = W_{C_t(\omega)}(\omega)$ must remain constant over the same interval. This observation proves the following.

7.28 PROPOSITION. *Let X be a continuous local martingale. Suppose that it is also a Stieltjes process. Then, for almost every ω, we have $X_t(\omega) = X_0(\omega)$ for all $t \geq 0$.* □

To put it another way, if X is a continuous local martingale and shows some signs of life, then its total variation must be infinite over some intervals.

Notes and Comments

The aim here is to provide some background and references for the material covered and for some extensions. No attempt is made to give a systematic account of all the sources.

Chapter I

This is a minimalist review of measure and integration for the probabilist. The analysts treat sigma-algebras and measurability as technical matters to be endured. The probabilists are serious: to them, a sigma-algebra represents a body of information, and measurability means observability and determinability. Consequently, most books on probability include at least an appendix on abstract measure theory. In particular, Billingsley (1979) and Dudley (1989) treat measure and probability together, as one topic. Doob (1994) is interesting along the same lines.

The classical text on measure theory is Halmos (1950). Royden (1968) and Rudin (1976) are also useful classics. The best reference for measure theory is Cohn (1980); it has a straightforward treatment which is well-informed of the needs of probability theory.

The notational principles used throughout are those of Dellacherie and Meyer (1975–87). For the integral notation μf, for instance, we follow the maxim that the most useful concepts should have the simplest notations. When integral signs appear, we follow the physicists and put the measure element next to the integral sign. This follows the summation notation and is almost necessary when there are multiple integral signs. Even the most tradition bound probabilist will use this system when several Markov kernels appear on the same line.

Standard measurable spaces are also called Borel spaces; for a brief introduction to them and to related issues, see the appendices in Dynkin and Yushkevich (1979).

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1,

Chapter II

The material in this chapter is standard ever since Kolmogorov (1933). It can be found in most books on probability. Billingsley (1979), Dudley (1989), Durrett (1995), Breiman (1968), Shiryaev (1995), Chung (1974), and Neveu (1965) are the main examples. The last two have been the main influences on my thinking. The comprehensive work Kallenberg (2002) is the best reference for researchers.

The notations and terminology are standard, except for the term "independency" for a collection of mutually (totally) independent things. The distribution functions are avoided; as Neveu (1965) noted, they should have disappeared long ago.

Chapter III

The material here is available in most textbooks. The treatment parallels Chung (1974) and Lamperti (1966). Most of the results belong to the classic period of probability, the first half of the 20th century. The clever proof of the strong law is due to Etemadi (1981). For further reading on characteristic functions and convergence in distribution, see Feller (1971) and Loève (1977). For weak convergence on general spaces, we refer to Billingsley (1968), Dudley (1989), and Pollard (1984).

Chapter IV

Conditional expectations are introduced heuristically first and are explained as projections in L^2-spaces and also through the Radon-Nikodym theorem. The concept is fundamental in probability theory. There are many related concepts: conditional probabilities, conditional distributions, conditional densities, etc. We made an effort to show the connections and to unify the notational systems. Much of the material here is standard; we follow Neveu (1965).

Conditional distributions are often the primary data in applications. Constructions of probability spaces start with them. The construction in the discrete-time case illustrates the point well; it is due to Cassius Ionescu-Tulcea (1949). A number of special cases are given to illustrate "stochastic modeling" of periodicity and temporal inhomogeneity in applications.

Chapter V

Filtrations and stopping times are introduced along standard lines, as in Dellacherie (1972) and in Dellacherie and Meyer (1975–87). The notation $\mathbb{E}_t X$ for the conditional expectation of X given $\mathcal{F}_t$ goes back to Lévy (1936) where martingales appear for the first time.

The term "martingale" is used first by Ville (1939). The theory is mostly due to Doob, who recognized its intrinsic importance and made it a versatile tool in probability theory. The modern theory is covered well in Doob (1953), Neveu (1975), Dellacherie and Meyer (1975–87). The account here is influenced by these and Meyer (1972) and the excellent exposition in Williams (1991).

The integral in 3.4 was first introduced in Burkholder (1966). The martingale convergence theorem is Doob's; its basic ingredient is Doob's upcrossing inequality, which is proved here using the financial intuition of Steele (2001). I followed the past literature and attributed Theorem 4.2 to Hunt; Blackwell and Dubins (1962) is earlier and has the same result with a slight generalization. For strong inequalities, the classic reference is Burkholder, Davis, Gundy (1972).

In continuous time, the treatment here is organized around Doob's stopping theorem; this is similar to making the strong Markov property the central concept for Markov processes. The term "Doob martingale" is coined for the central object; it is a stopped martingale that is uniformly integrable, that is, a stopped martingale of "class D". This treatment seems to achieve some simplicity both in theory and in applicability.

Martingale characterization for Brownian motion is due to Lévy (1948); we followed Doob (1953) for its proof, a bare hands approach, without using the modern tools of stochastic calculus. The martingale characterization for Poisson processes is due to Watanabe (1964); for some extensions see Brémaud (1981).

For further treatments on martingales, starting points are, in addition to the books mentioned above, Jacod (1979), Jacod and Shiryaev (1987), Durrett (1984), Revuz and Yor (1991), and most books on stochastic calculus.

Chapter VI

Poisson random measures have manifold uses in statistics, in applied probability, and in the theories of point processes and Markov processes. In chapters to follow, they will be used to explain the structure of Lévy processes, the excursions of Brownian motion, and the jumps of Markov processes. This chapter is to give a serious introduction to them.

Kallenberg (1983) is a rigorous introduction to random measures and weak convergence of such. Many related matters, extensions, and applications can be found in Jacod (1975), Karr (1991), Kingman (1993), Serfozo (1999), Jacobsen (2006), and Daley and Vere-Jones (2008).

The main theorem on random transformations is Theorem 3.2. Its origins seem lost in time. It is called the marking theorem in Kingman (1993); the version in Theorem 3.26 is due to Serfozo (1999) with a missing condition. A forerunner of it, the application mentioned following 3.13, appears in Doob (1953) with a long proof.

Additive random measures are generalizations, to abstract spaces, of additive processes of Lévy (1948) and Itô (1942); see Itô (2004) for a complete account of such processes. On abstract spaces, they are called completely random measures in Kingman (1967), (1993); the converse mentioned immediately after Theorem 4.4 is due to Kingman (1967).

Section 5 brings together many strands on Poisson processes considered as submartingales, as Markov processes, as Lévy processes, and as point processes. The simplest characterization is in Theorem 5.9; the proof here follows Doob (1953) with a slight improvement. Theorem 5.12 is due to Kallenberg.

Section 6 is an introduction to stochastic calculus driven by Poisson random measures and the uses of it to construct self-exciting point processes. Similar material can be found in Brémaud (1981), Jacod (1979), Jacod and Shiryaev (1987).

Chapter VII

Lévy processes are the continuous time versions of partial sums of independent and identically distributed random variables. Much of the pioneering work is summarized in Lévy (1937). The stochastic approach we followed is due to Itô (1942), where the first sentence is "The aim of this paper is to give a new rigorous proof of the known formula of P. Lévy (1937) on the infinitely divisible law (of probability) by making use of the scheme of stochastic differential processes introduced by J.L. Doob (1937)". Such processes were called variously as additive processes, differential processes, or processes with stationary and independent increments. The term "Lévy process" was coined by P.-A. Meyer.

There are at least two good books on them: Bertoin (1996) treats them as special Markov processes; Sato (1999) has wider coverage, especially of connections to infinite divisibility. The treatment here follows Itô (2004) with some assist from Sato (1999). The strong Markov property, Theorem 3.10, was historically the first rigorous formulation of that property; it is due to Hunt (1956). The difficult case mentioned in Remark 7.10 was proved by Kesten (1969) for Lévy processes in d-dimensions; see Bretagnolle (1971) for the same. The problem is important in the theory of Markov processes. Indeed, subordinators are important in the general theory of regenerative systems; see Maisonneuve (1971) for the connections.

Lévy processes have stationary and independent increments. Dropping stationarity, we get additive processes of Itô and Lévy. For them we refer to Sato (1999), Itô (2004), and Jacod (1979). In fact, increasing additive processes were treated in Chapter VI, Section 4, under the name of additive measures; take $E = \mathbb{R}$ there. A further generalization is to Markov additive processes introduced in Çınlar (1972). Such a process is a pair (X,Y) where X is a Markov process and the conditional law of Y given X is that of an

additive process whose local characteristics are modulated by X. They found applications to regenerative processes and excursions of Markov processes from a set of states; see Çınlar (1975), Maisonneuve (1974), Kaspi (1983), and Çınlar and Kaspi (1983).

Lévy, additive, and Markov additive processes are used heavily in applied probability. Moran (1956) was the first to use gamma processes to model the input process to a dam; see Çınlar (1973) for another instance. See Breuer (2003) and Pachego, Tang, Prabhu (2009) for uses in queueing and telecommunications. An area of intense applications is mathematical finance; see Applebaum (2009), Cont (2003), and Schoutens (2003) for starters.

Chapter VIII

This introduction to Brownian motion is along the lines of Lévy (1948) and Itô and McKean (1965). For more comprehensive accounts see Freedman (1971), Karatzas and Shreve (1988), Revuz and Yor (1991), and Mörters and Peres (2010). The present treatment differs from the earlier ones by the methodical use of Poisson random measures, especially on matters related to excursions and local times.

Hitting times, maximum processes, and local times were the brainchildren of Lévy's, along with much else. For most of these matters the complete reference is Revuz and Yor (1991). Theorem 4.19 is due to Pitman (1975) with a proof via random walk approximations. See Williams (1979) for a great exposition of all these. The source on the Poisson nature of excursions is Theorem 6.14, which is due to Itô (1970). In identifying the Itô measure of excursions, we chose to disintegrate it with respect to the Lévy measure of extents; this is due to Williams and is explained in Williams (1979); we followed the proof in Rogers (1981) with minor changes. Another approach is that in Ikeda and Watanabe (1989), where the Itô measure is disintegrated with respect to the Lévy measure of durations. For further appearances of Poisson random measures in Brownian theory see Pitman (1981) and Çınlar (1992).

For path properties of Section 7 see Freedman (1971), Knight (1981), Karatzas and Shreve (1988), and Mörters and Peres (2010). The construction of Theorem 8.6, due to Lévy, is the best way to visualize the paths. Finally, a fantastic collection of facts and results is available in Borodin and Salminen (1996).

The titubations of a microscopic particle suspended in water were first observed by the botanist Robert Brown in 1826 and were the cause of much excitement in religio-scientific circles. Unaware of the phenomenon, Einstein predicted it in 1905 by theoretical reasoning as the visible consequence of the unseen molecular motions of the surrounding medium; he derived the heat equation for the visible motion, cW in our terms, and gave a formula for the constant c in terms of physical quantities. For us the theory starts with

Bachelier (1900) and with Wiener (1920), the latter showing the existence of the mathematical Brownian motion via construction as a random Fourier series.

Chapter IX

Markov processes form the most extensively studied part of probability theory. They are ubiquitous in the theory and applications of probability. Indeed, as clarified in 1.12, it is an old remark of Doob's that every process becomes Markovian if one calls "state at time t" the whole past until t. An even better idea is found in Knight (1992): given a process, one obtains a good Markov process if one calls "state at time t" the conditional law of the future given the past until t.

We use the term "chain" only for processes with discrete time-parameter. The best reference to their theory is Revuz (1984). The special case of random walks is of great importance; see Lawler and Limic (2010) for them. For chains with discrete state spaces, almost every book on applied probability gives a treatment of ergodic behavior following the treatment in Feller (1957b); for an alternative approach see Thorison (2000); and for the same for arbitrary state spaces see Nummelin (2004) and Meyn and Tweedy (2009). In the case of transient chains with discrete state spaces, the boundary theory is explored in Feller (1956), Hunt (1960), and Dynkin (1969). For the potential theory of chains see Revuz (1984) and Kemeny, Snell, Knapp (1966), the latter being a comprehensive treatment for chains with discrete state spaces.

We concentrate on Markov processes (with continuous time-parameter). Their theory starts with Kolmogorov (1931); introduced there are the notions of Markov kernels and transition functions, the Chapman-Kolmogorov equation, Kolmogorov's differential equations in the discrete state space case, and the partial differential equations corresponding to diffusions. Much of the early theory followed Kolmogorov and was concerned with transition functions, evolution equations for them, and their ergodic behavior. This period culminates with Feller (1954a), (1954b), (1955) on the characterization of differential operators that are generators for diffusions on the real line. An account can be found in Karlin and Taylor (1981) and also in Dynkin (1965).

The modern theory starts with Itô (1942), where the dynamics of motion is described by stochastic integral equations, and transition functions and generators become derivative concepts. As developed in Itô (1942) and (1951), the Markovian motion is an integral path in a field of Lévy processes: when the particle is at x, the tangent to its path is a Lévy process whose characteristics depend on x. This is the meaning of the equation 3.95 and its particular cases 2.2 and 3.1. We concentrate on the particular cases in Sections 2 and 3. For the needs of Section 2, we give a brief introduction to stochastic integration in Section 7 following he formalism of Ikeda and Watanabe (1989); for further reading we refer to McKean (1969), Letta (1984), Chung and Williams (1990), Øksendal (2003), and Protter (2003).

Section 2 introduces Itô diffusions. The equation of motion is a stochastic differential equation driven by Wiener processes. Under a Lipschitz condition on the coefficients, it has a unique strong solution, which is a continuous strong Markov process (a diffusion, in short). The interplays between the generators, resolvents, and transition functions are derived easily. For diffusions in general, see Itô and McKean (1965). For a deep study of the equation of motion in one-dimension, we refer to Engelbert and Schmidt (1989–91); see Assing and Schmidt (1998) also. Generators establish a link between Markov processes and partial differential equations; see Dynkin (1982) for connections to some problems of classical analysis. Since a Markov process is a much richer structure than its generator or its transition function, the probability theory can obtain analytical results which are beyond the analytic theory; see Stroock and Varadhan (1979) for such examples; see Williams (1974) for an excellent expository account.

Theorem 2.23 is the key relationship between generators and martingales. More generally, let G be defined by 2.21 but without the Lipshitz condition on the coefficients; if 2.24 defines a martingale for every f there, then X is a Markov process. This is the martingale characterization for Markov processes; it was introduced in Stroock and Varadhan (1969) and elaborated on in Stroock and Varadhan (1979). See Ethier and Kurtz (1986) for the same and for generators of semigroups in general. See Jacod (1979) for martingale problems in more general settings.

Section 3 introduces Itô processes with jumps. The equation of motion is a stochastic integral equation driven by Wiener processes and a standard Poisson random measure. At first, we work under conditions that make the motion piecewise continuous. The generators, resolvents, and transition functions are derived as secondary objects and are related to the diffusions between the successive jumps. The theory here is a blend of that for diffusions and Poisson random measures. It can be extended to the more general case where 3.95 is the equation of motion; the general case allows infinitely many jumps during small intervals of time; Stroock (1975) gives sufficient conditions that yield strong Markov processes.

Later half of Section 3 is devoted to the case where there is no movement between the jumps, and the jump times are allowed to have accumulation points. In part, this is a modern introduction (using integral equations and Poisson random measures) to processes with discrete state spaces; they were studied deeply in Chung (1960), (1963), (1966), (1970), and Feller (1957a).

Section 4 introduces Markov processes in the abstract setting pioneered by Hunt (1957–58). The Markov process is introduced axiomatically and takes the center stage; generators are eliminated, and resolvents and transition semigroups are secondary objects defined by the process. Here, and in Section 5 and 6, we follow the classic text Blumenthal and Getoor (1968). Section 5 introduces Hunt processes; these are Markov processes of Section 4 with the added assumptions of strong Markov property and quasi-left-continuity. We discuss the last two properties and their implications, and we show that

the Feller property for the transition semigroup implies that the process is Hunt. In discussing the stopping times, we replaced the standard terms "accessible" and "totally inaccessible" with the terms "sigma-predictable" and "totally unpredictable"; the rectified terms seem more natural and more correct; the earlier terms allude to the hitting times of accessible and totally inaccessible subsets of the state space. For the general theory of Markov processes, the best references are Sharpe (1988) and the volumes 4 and 5 of Dellacherie and Meyer (1975–87). For excursions of Markov processes, generalizing Itô's excursion theory for Brownian motions, see Blumenthal (1992) and Kaspi (1983). Section 6 is an introduction to Hunt's extension of the classical potential theory. For this material, the best treatments are in the works mentioned above and Chung and Walsh (2005).

Finally, what is a Hunt process? The answer shows that Wiener and Poisson processes are the building blocks for all Hunt processes. Let Y be a Hunt process whose state space is $\mathbb{R}^d$ and suppose that each one of the d components is a semimartingale. Then there exist a Wiener process on $\mathbb{R}^d$ and a standard Poisson random measure such that, after a random time change, Y becomes a process X that satisfies the equation 3.95 for some choice of the coefficients a, b, j. Thus, the only randomness in Y comes from Wiener and Poisson processes. We refer to Çınlar and Jacod (1981) – there, the comment in page 164 on continuous processes with bounded variation turns out to be too hasty. Preliminaries for this paper and the connections between martingale theory and Markov processes can be found in Çınlar, Jacod, Protter, and Sharpe (1980). The case of a Hunt process Z with a Lusin state space E is somewhat similar: one can find a continuous function $\mathrm{f}: \mathrm{E} \mapsto \mathbb{R}^\infty$ such that $\mathrm{f} \circ \mathrm{Z}$ is like the process Y above, but infinite dimensional.

Bibliography

D. APPLEBAUM (2009) *Lévy Processes and Stochastic Calculus*, 2nd ed. Cambridge Univ. Press, Cambridge.

L. ARNOLD (1974) *Stochastic Differential Equations.* Wiley, New York.

S. ASSING, W.M. SCHMIDT (1998) Continuous Strong Markov Processes in Dimension One. *Lecture Notes Math.* **1688**. Springer-Verlag, Berlin.

L. BACHELIER (1900) Théorie de la spéculation. *Ann. Sci. l'École Normale Super. Ser 3*, **17**, 21–86. (Reprinted in 1995 by Gabay, Paris.)

J. BERTOIN (1996) *Lévy Processes.* Cambridge Univ. Press, Cambridge.

P. BILLINGSLEY (1968) *Convergence of Probability Measures.* Wiley, New York.

P. BILLINGSLEY (1979) *Probability and Measure.* Wiley, New York.

D. BLACKWELL and L. DUBINS (1962) Merging of opinions with increasing information. *Ann. Math. Statist.* **33**, 882–886.

R.M. BLUMENTHAL (1992) *Excursions of Markov Processes.* Birkhäuser, Boston.

R.M. BLUMENTHAL and R.K. GETOOR (1968) *Markov Processes and Potential Theory.* Academic Press, New York. (Reprinted in 2007 by Dover, Mineola, NY.)

A.N. BORODIN, P. SALMINEN (1996) *Handbook of Brownian Motion – Facts and Formulae.* Birkhäuser, Basel.

L. BREIMAN (1968) *Probability.* Addison-Wesley, Reading, MA.

P. BREMAUD (1981) *Point Processes and Queues.* Springer, New York.

J. BRETAGNOLLE (1971) Résultats de KESTEN sur les processus à accroissements indépendants. *Séminaire de Probabilités* V, pp. 21–36. *Lecture Notes Math.* **191**. Springer-Verlag, Berlin.

L. BREUER (2003) *From Markov Processes to Spatial Queues*. Springer, Berlin.

D.L. BURKHOLDER (1966) Martingale transforms. *Annals of Math. Statist.* **37**, 1494–1504.

D.L. BURKHOLDER, B.J. DAVIS, R.F. GUNDY (1972) Integral inequalities for convex functions of operators on martingales. *Proc. 6th Berkeley Symposium Math. Statist. Probab. 2*, 223–240. Univ. California Press, Berkeley.

K.L. CHUNG (1960) *Markov Chains with Stationary Transition Probabilities*. Springer, Berlin.

K.L. CHUNG (1963) On the boundary theory for Markov chains. *Acta Math.* **110**, 19–77.

K.L. CHUNG (1966) On the boundary theory for Markov chains. *Acta Math.* **115**, 111–163.

K.L. CHUNG (1970) *Lectures on Boundary Theory for Markov Chains*. Princeton Univ. Press, Princeton.

K.L. CHUNG (1974) *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

K.L. CHUNG and J.B. WALSH (2005) *Markov Processes, Brownian Motion, and Time Symmetry*, 2nd ed. Springer, New York.

K.L. CHUNG and R.J. WILLIAMS (1990) *Introduction to Stochastic Integration*. Birkhäuser, Boston.

E. ÇINLAR (1972) Markov additive processes, I and II. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **24**, 85–93 and 94–121.

E. ÇINLAR (1973) Theory of continuous storage with Markov additive inputs and a general release rule. *J. Math. Anal. Appl.* **43**, 207–231.

E. ÇINLAR (1975) Lévy systems of Markov additive processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **31**, 175–185.

E. ÇINLAR (1992) Sunset over Brownistan. *Stochastic Process Appl.* **40**, 45–53.

E. ÇINLAR and J. JACOD (1981) Representation of semimartingale Markov processes in terms of Wiener processes and Poisson random measures. *Seminar on Stochastic Processes 1981*, pp. 159–242. Birkhäuser, Boston.

E. ÇINLAR, J. JACOD, P. PROTTER, M.J. SHARPE (1980) Semimartingales and Markov processes. Z. W*ahrscheinlichkeitstheorie verw. Gebiete* **54**, 161–219.

E. ÇINLAR and H. KASPI (1983) Regenerative systems and Markov additive processes. *Seminar on Stochastic Processes 1982*, pp. 123–147. Birkhäuser, Boston.

D.L. COHN (1980) *Measure Theory.* Birkhäuser, Boston.

R. CONT (2003) *Financial Modeling with Jump Processes.* Chapman & Hall, Boca Raton.

D.J. DALEY and D. VERE-JONES (2008) *An Introduction to the Theory of Point Processes*, vol.2, 2nd ed. Springer, New York.

C. DELLACHERIE (1972) *Capacités et Processus Stochastiques.* Springer, Berlin.

C. DELLACHERIE, B. MAISONNEUVE, P.-A. MEYER (1992) *Probabilités et Potentiel*, vol. *V*. Hermann, Paris.

C. DELLACHERIE, P.-A. MEYER (1975–1987) *Probabilités et Potentiel*, vol*s.I–IV*. Hermann, Paris.

J.L. DOOB (1937) Stochastic processes depending on a continuous parameter. *Trans. Amer. Math. Soc.* **42**, 107–140.

J.L. DOOB (1953) *Stochastic Processes.* Wiley, New York.

J.L. DOOB (1957) Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France*, **85**, 431–458.

J.L. DOOB (1958) Probability and the first boundary value problem. *Illinois J. Math.* **2**, 19–36.

J.L. DOOB (1994) *Measure Theory.* Springer, New York.

R.M. DUDLEY (1989) *Real analysis and Probability.* Wadsworth, Belmont, CA.

R. DURRETT (1984) *Brownian Motion and Martingales in Analysis.* Wadsworth, Belmont, CA.

R. DURRETT (1995) *Probability: Theory and Examples.* Wadsworth, Pacific Grove, CA.

E.B. DYNKIN (1965) *Markov Processes.* Academic Press, New York.

E.B. DYNKIN (1969) Boundary theory of Markov processes (the discrete case). *Russ. Math. Surveys*, **24**, 1.

E.B. DYNKIN (1982) *Markov Processes and Related Problems of Analysis.* Cambridge Univ. Press, Cambridge.

E.B. DYNKIN and A.A. YUSHKEVICH (1979) *Controlled Markov Processes.* Springer, Berlin.

H.J. ENGELBERT, W. SCHMIDT (1989–1991) Strong Markov continuous local martingales and solutions of one-dimensional stochastic differential equations I, II, III. *Math. Nachr.* **143**, 167–184; **144**, 241–281; **151**, 149–197.

N. ETEMADI (1981) An elementary proof of the strong law of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **55**, 119–122.

S.N. ETHIER and T.G. KURTZ (1986) *Markov Processes: Characterization and Convergence.* Wiley, New York.

W. FELLER (1954a) Diffusion processes in one dimension. *Trans. Amer. Math. Soc.* **77**, 1–31.

W. FELLER (1954b) The general diffusion operator and positivity preserving semi-groups in one dimension. *Ann. Math.* **60**, 417–436.

W. FELLER (1955) On second order differential operators. *Ann. Math.* **61**, 90–105.

W. FELLER (1956) Boundaries induced by non-negative matrices. *Trans. Amer. Math. Soc.* **83**, 19–54.

W. FELLER (1957a) On boundaries and lateral conditions for the Kolmogorov differential equations. *Ann. Math.* **65**, 527–570.

W. FELLER (1957b) *An Introduction to Probability Theory and Its Applications*, vol.1, 2nd ed. Wiley, New York.

W. FELLER (1971) *An Introduction to Probability Theory and Its Applications*, vol.2, 2nd ed. Wiley, New York.

D. FREEDMAN (1971) *Brownian Motion and Diffusion.* Holden-Day, San Francisco.

D. FREEDMAN (1983) *Markov Chains.* Springer-Verlag, New York. (Original published in 1971 by Holden-Day, San Francisco).

G.R. GRIMMETT and D.R. STIRZAKER (2001) *Probability and Random Processes*, 3rd ed. Oxford Univ. Press, Oxford.

P.R. HALMOS (1950) *Measure Theory.* Van Nostrand, Princeton.

G.A. HUNT (1956) Some theorems concerning Brownian motion. *Trans. Amer. Math. Soc.* **81**, 294–319.

G.A. HUNT (1957–58) Markoff processes and potentials **I**, II, and III. *Illinois J. Math.* **1**, 44–93; **1**, 316–369; and **2**, 151–213.

G.A. HUNT (1960) Markoff chains and Martin boundaries. *Illinois J. Math.* **4**, 313–340.

N. IKEDA and S. WATANABE (1989) *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. North Holland, Amsterdam.

C. IONESCU-TULCEA (1949) Mesures dans les espaces produits. *Atti. Acad.Naz. Lincei Rend.* **7**, 208–211.

K. ITÔ (1942) On stochastic processes, I: infinitely divisible laws of probability. *Japan J. Math.* **18**, 261–301.

K. ITÔ (1951) On stochastic differential equations. *Memoirs Amer. Math. Soc.* **4**.

K. ITÔ (1970) Poisson point processes attached to Markov processes. *Proc. 6th Berkeley Symposium Math. Statist. Probab.* vol.3, 225–239. Univ. California Press, Berkeley.

K. ITÔ (1987) *Selected Papers*, ed. by D.W. Stroock and S.R.S. Varadhan. Springer-Verlag, New York.

K. ITÔ (2004) *Stochastic Processes.* Springer, Berlin.

K. ITÔ and H.P. McKEAN (1965) *Diffusion Processes and Their Sample Paths.* Springer-Verlag, Berlin.

M. JACOBSEN (2006) *Point Process Theory and Applications.* Birkhäuser, Boston.

J. JACOD (1975) Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **31**, 235–253.

J. JACOD (1979) *Calcul Stochastique et Problèmes de Martingales.* Lecture Notes Math. **714**. Springer-Verlag, Berlin.

J. JACOD, A.N. SHIRYAEV (1987) *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

O. KALLENBERG (1983) *Random Measures*, 3rd ed. Academie-Verlag, Berlin.

O. KALLENBERG (2002) *Foundations of Modern Probability*, 2nd ed. Springer, New York.

I. KARATZAS, S.E. SHREVE (1988) *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

S. KARLIN and H. TAYLOR (1975) *A First Course in Stochastic Processes*, 2nd ed. Academic Press, New York.

S. KARLIN and H. TAYLOR (1981) *A Second course in Stochastic Processes.* Academic Press, New York.

A.F. KARR (1991) *Point Processes and Their Statistical Inference*, 2nd ed. Marcel Dekker, New York.

H. KASPI (1983) Excursions of Markov processes: An approach via Markov additive processes. Z. W*ahrscheinlichkeitstheorie verw. Gebiete* **64**, 251–268.

J.G. KEMENY, J.L. SNELL, A.W. KNAPP (1966) *Denumerable Markov Chains.* Van Nostrand, Princeton.

H. KESTEN (1969) Hitting probabilities for processes with stationary independent increments. *Memoirs Amer. Math. Soc.* **93**.

J.F.C. KINGMAN (1967) Completely random measures. *Pacific J. Math.* **21**, 59–78.

J.F.C. KINGMAN (1993) *Poisson Processes.* Oxford Univ. Press, Oxford.

F. KNIGHT (1981) *Essentials of Brownian Motion and Diffusions.* Amer. Math. Soc. Providence, RI.

F. KNIGHT (1992) *Foundations of the Prediction Process.* Oxford Univ. Press, Oxford.

A.N. KOLMOGOROV (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* **104**, 415–458.

A.N. KOLMOGOROV (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin.

J. LAMPERTI (1966) *Probability.* Benjamin, New York.

G.F. LAWLER and V. LIMIC (2010) *Random Walk.* Cambridge Univ. Press, Cambridge.

G. LETTA (1984) Martingales et Intégration Stochastique. Scuola Normale Superiore, Pisa.

P. LÉVY (1936) La loi forte des grands nombres pour les variables aléatoires enchaînées. *J. Math. Pures et Appl.* **15**, 11–24.

P. LÉVY (1937) *Théorie de l'Addition des Variables Aléatoires.* Gauthier-Villars, Paris.

P. LÉVY (1948) *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, Paris; (2nd ed. 1965 is reprinted by Gabay, Paris, in 1995).

M. LOÈVE (1977) *Probability Theory*, 4th ed. Springer, New York.

B. MAISONNEUVE (1971) Ensembles régénératifs, temps locaux et subordinateurs. *Séminaire de Probabilités* V, pp. 147–169. *Lecture Notes Math.* **191**. Springer-Verlag, Berlin.

B. MAISONNEUVE (1974) Systèmes régénératifs. *Astérisque* **15**, Soc. Math. France.

H.P. McKEAN (1969) *Stochastic Integrals.* Academic Press, New York.

P.-A. MEYER (1972) Martingales and Stochastic Integrals I. *Lecture Notes Math.* **284**. Springer-Verlag, Berlin.

S. MEYN and R.L. TWEEDIE (2009) *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge.

P.A.P. MORAN (1956) A probability theory of a dam with a continuous release rule. *Quart. J. Math. Oxford* 7, 130–137.

P.A.P. MORAN (1959) *The Theory of Storage.* Methuen, London.

P. MÖRTERS and Y. PERES (2010) *Brownian Motion.* Cambridge University Press, Cambridge.

J. NEVEU (1965) *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

J. NEVEU (1975) *Discrete Parameter Martingales.* North-Holland, Amsterdam.

E. NUMMELIN (2004) *General Irreducible Markov Chains and Non-negative Operators.* Cambridge Univ. Press. Cambridge.

B. ØKSENDAL (2003) *Stochastic Differential Equations*, 6th ed. Springer-Verlag, Berlin.

A. PACHECO, L.C. TANG, N.U. PRABHU (2009) *Markov-Modulated Processes and Semiregenerative Phenomena.* World Scientific, Singapore.

J.W. PITMAN (1975) One-dimendional Brownian motion and the three-dimensional Bessel process. *J. Appl. Probab.* 7, 511–526.

J.W. PITMAN (1983) Remarks on the convex minorant of Brownian motion. *Seminar on Stochastic Processes 1982*, pp. 219–227. Birkhäuser, Boston.

D. POLLARD (1984) *Convergence of Stochastic Processes.* Springer, New York.

P. PROTTER (2003) *Stochastic Integration and Differential Equations*, 2nd ed. Springer, New York.

D. REVUZ (1984) *Markov Chains*, 2nd ed. Elsevier, Amsterdam.

D. REVUZ, M. YOR (1991) *Continuous Martingales and Brownian Motion.* Springer, Berlin.

L.C.G. ROGERS (1981) William's characterization of the Brownian excursion law: proof and applications. *Séminaire de Probabilités* XV, pp. 227–250. *Lecture Notes Math.* **850**. Springer-Verlag, Berlin.

L.C.G. ROGERS and D. WILLIAMS (1987) *Diffusions, Markov Processes, and Martingales*, vol. 2. Wiley, New York.

H.L. ROYDEN (1968) *Real Analysis*, 2nd ed. MacMillan, New York.

W. RUDIN (1976) *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, New York.

K. SATO (1999) *Lévy Processes and Infinitely Divisible Distributions.* Cambridge Univ. Press, Cambridge.

W. SCHOUTENS (2003) *Lévy Processes in Finance.* Wiley, New York.

R.F. SERFOZO (1999) *Introduction to Stochastic Networks.* Springer, New York.

M. SHARPE (1988) *General Theory of Markov Processes.* Academic Press, Boston.

A.N. SHIRYAEV (1995) *Probability*, 2nd ed. Springer, New York.

J.M. STEELE (2001) *Stochastic Calculus and Financial Applications.* Springer, New York.

D.W. STROOCK (1975) Diffusion processes associated with Lévy generators. Z. *Wahrscheinlichkeitstheorie verw. Gebiete* **32**, 209–244.

D.W. STROOCK and S.R.S. VARADHAN (1969) Diffusion processes with continuous coefficients I and II. *Comm. Pure Appl. Math.* **22**, 345–400 and 479–530.

D.W. STROOCK and S.R.S. VARADHAN (1979) *Multidimensional Diffusion Processes.* Springer-Verlag, Berlin.

H. THORISON (2000) *Coupling, Stationarity, and Regeneration.* Springer, New York.

J. VILLE (1939) *Étude critique de la notion de collectif.* Gauthier-Villars, Paris.

S. WATANABE (1964) On discontinuous additive functionals and Lévy measures of a Markov process. *J. Math. Soc. Japan* **16**, 83–95.

N. WIENER (1923) Differential space. *J. Math. Phys.* **2**, 131–174.

D. WILLIAMS (1974) Brownian motions and diffusions as Markov processes. *Bull. London Math. Soc.* **6**, 257–303.

D. WILLIAMS (1979) *Diffusions, Markov Processes, and Martingales*, vol.1. Wiley, New York.

D. WILLIAMS (1991) *Probability with Martingales*. Cambridge Univ. Press, Cambridge.

Index

A

Absolute continuity, 31, 34, 35, 370–371
Adaptedness, 172
Additive functionals, 340, 342, 516
Additive processes, 340, 516
Additive random measures, 277–290
Algebra, 2–8
Almost everywhere, 17, 500
Almost sure events, 50, 500
Arcsine distribution, 377, 393–395, 419, 424, 425
Arcsine laws, 393–395, 419, 424, 425
Arrival processes, 82, 173, 180, 258, 261, 262, 301, 309

B

Bernoulli sequences, 89–92
Bessel processes, 225, 379, 386–389, 399, 404–407, 418, 471–472
Blumenthal's zero-one law, 347, 354, 383, 390, 502
Borel-Cantelli lemmas, 98–99, 205–206
Branching processes, 304–305, 311
Brownian bridge, 388
Brownian motion
  excursions, 379, 413–422, 540
  geometric, 388, 456–457, 472
  hitting times, 220, 389–396, 506
  holding boundaries, 504
  maximum process, 396
  quantiles, 425
  recurrence, 389–396, 471
  skew, 422
  on the unit circle, 457, 470

C

Campbell's theorem, 253
Cantor set, function, 35–36, 408
Caratheodory extension theorem, 19, 163
Cauchy criterion, 94, 97, 100, 104
Cauchy distribution, 65, 330, 333, 334, 336, 339, 393, 394
Cauchy processes, 330, 332–335, 345, 363, 399, 416
Chapman-Kolmogorov equation, 170, 446, 447
Characteristic exponent, 315, 316, 318–320, 323, 329, 331, 337, 338, 355
Characteristic functions, 61, 67–69, 88, 116, 136, 315
Compensated sum of jumps, 322, 334
Compound Poisson process, 265–266, 287, 314–315, 319, 321, 351–354, 356, 364, 376, 502
Conditional distributions, 151–157
Conditional expectations
  given an event, 140, 149
  given a random variable, 146
  given a sigma-algebra, 140–142
  notations for, 147–148
Conditional independence, 158–160
Conditional probabilities, 148–157
  regular versions, 150–151
Convergence
  almost sure, 97–102
  in distribution, 110–118, 127–137, 263

of Fourier transforms, 115–116
in L^p, 105–109, 211
in probability, 101–107, 118, 121
of quantiles, 112–113
of real sequences, 93–97
of series, 124–127
Convolution, 85, 89
Counting processes, 173, 188, 230, 290–293, 301–304, 308–311, 349
Cylinder sets, 165

D
De Moivre-Laplace theorem, 117, 127
Determinability, 75–82, 533
Dirichlet random measure, 288
Disintegration, 153–154
Disjointedness, 12
Distribution functions, 33, 36, 51, 55–57, 96, 112–113, 122, 534
Distributions
arcsine, 377, 393–395, 419, 421, 424, 425
beta, 62, 288, 377, 393, 421, 424, 425
binomial, 90, 134, 295
Cauchy, 65, 330, 333, 334, 336, 339, 393, 394
Dirichlet, 288
exponential, 54, 55, 62, 65, 148, 224, 258, 277, 291, 297, 354, 395, 450, 464
gamma, 54, 55, 61–66, 89, 134, 281, 282, 287, 319, 327, 339, 425
Gaussian, 54, 64, 66, 68, 69, 89, 117, 127, 129–134, 156, 186, 232, 314, 337, 338, 350, 351, 380, 382, 384, 388, 389, 439, 448
geometric, 90
Linnik, 368
Multinomial, 91, 259
Pareto, 63
Poisson, 53, 89, 117, 123, 132–134, 188, 230, 245, 248–252, 257, 291, 349
symmetric gamma, 327
uniform, 56, 65, 276, 288, 332, 335, 337, 338, 363, 367, 393, 424
Dominated convergence, 25–27, 29, 59
for conditional expectation, 143, 205
Doob-Kolmogorov inequality, 197
Doob martingales, 214–220, 222, 223, 462, 471, 520, 535
Doob's
decomposition, 190–191, 198, 199, 206, 211, 308, 519
maximal inequalities, 223
norm inequality, 197, 357
stopping theorem, 193–194, 202–203, 214, 216–217
Dynkin's formula, 461–462, 470, 471

E
Empirical distribution, 91, 122
Excessive functions, 444, 518–524
Excursions, 413–426, 535, 537, 540
extent of, 415–417
Exponential martingales, 187, 224, 457

F
Fatou's lemma, 25, 59, 117
for conditional expectations, 143
Feller-Levy theorem, 132
Feller processes, 513–515
Feynman-Kac formula, 467, 481
Filtrations
augmentation, 234–236, 346
right-continuity, 235, 502
Fourier transforms, 60–61, 115–116, 316, 335, 363
Fubini's theorem, 42, 44
Functions
Borel, 7, 13, 21, 36

compositions of, 6–7
elementary, 12
indicators, 8–9
measurable, 6–13, 19, 26, 27
numerical, 7, 8, 10–12, 17, 19
positive parts, 7–8
simple, 8–10, 12, 23

G

Gamma
distribution, 54, 55, 61–66, 89, 134, 281
processes, 280–281, 285, 287, 315, 320, 345, 362, 367, 368, 375, 537
two-sided, 319, 327
random measures, 282, 288
Gaussian vectors, 68, 69
Gauss-Markov chains, 448
Generators, 460–463, 466, 478–479, 484, 485, 490, 493, 496, 538, 539
Geometric Brownian motion, 388, 456–457, 472
Glivenko-Cantelli, 122
Gronwall's inequality, 472–473

H

Harmonic functions, 185, 520
Hawkes processes, 311
Helly's theorem, 96, 115
Holder continuity, 426, 428–429
Holding points, 354, 502–504, 510–511
Hunt's dominated convergence, 205
Hunt system, 505, 514, 516–518

I

Increasing Levy processes, 279–284, 286, 313, 315, 317, 319, 360–364, 368–377, 397, 453
Increasing processes, 190, 226–228, 248, 355
Independence
of collections, 53, 82–84, 92
of increments, 279, 296
of random variables, 52–53, 82–91
of sigma-algebras, 82–87, 158
Indistinguishability, 241, 526
Inequalities
Chebyshev's, 64
Holder's, 70, 71
Jensen's, 64, 70–71, 143
Markov's, 64, 346
Minkowski's, 70, 71
Inequalities for maxima, 124–125
Doob's, 197, 223
Kolmogorov's, 124, 197
Infinite divisibility, 134, 315, 536
Infinite server queues, 274
Infinitesimal generator, 462, 463
Information, 75–82, 139, 140, 142, 144, 147, 158, 171–174, 178, 181–183, 190–193, 235, 533
Instantaneous points, 502–503
Integrability, 21–22, 28, 42, 142
uniform, 70–75, 106–107, 183–184, 189, 200–201, 204, 215, 217–218
Integrals
indefinite, 29–37
Lebesgue-Stieltjes, 36, 37, 191
Riemann, 21
Intensity process, 262, 301, 302, 308, 311, 466
Ionescu-Tulcea theorem, 162–165, 254
Ito diffusions, 443, 455–474, 497, 502, 505, 515, 539
Ito-Levy decomposition, 322, 354–360, 398, 497
Ito measure of excursions, 420
Ito's formula, 460–461, 470, 528–530

J

Jensen's inequality, 64, 70, 131, 143, 184
Jump measure, 358–359
Jump size, 316, 319, 327, 376

K
Kernels
  finite, sigma-finite, 37, 39–47
  identity, 39, 245
  Levy, 476, 482, 485–486, 494–496
  Markov, 39, 40, 167–169, 184, 185, 445–447, 453, 464, 466, 482, 499, 503, 533, 538
  products, 37–47
Kolmogorov
  extension theorem, 164–166, 169, 437, 452
  equations, 443, 463, 479
  inequality, 124, 125, 197, 324, 381
  three series theorem, 126–127
  zero-one law, 85–86, 99, 124, 199, 207, 347
Kronecker's lemma, 96–97, 121, 127

L
L^p-spaces, 70–75
  convergence in, 93
Langevin equation, 457
Laplace functional, 244–245, 247, 251–254, 258, 286
Laplace transforms, 61–63, 65–68, 88, 89, 123, 133, 220–222, 244, 368–370, 395, 399, 425, 463, 483, 484, 492
Law of the iterated logarithm, 430–434
Laws of large numbers, 93, 118–123
Levy-Khinchine formula, 315, 322–323, 355
Levy
  characterization of Wiener, 293–294
  measure, 136, 280, 287, 318, 327–336, 339, 354, 355, 358, 359, 361–370, 410, 416, 537
  processes
    characteristics for, 323, 355, 364
    increasing, 279–282, 286, 315, 317, 338, 360, 368–377
  property, 340, 342
Levy's increasing continuous process, 450
Liapunov's theorem, 129–130
Lindeberg's theorem, 131–132, 134
Lipschitz continuity, 456, 458, 479
Local martingales, 219–220, 223, 471, 472, 525, 527, 530–532
Local times, 407–416, 419, 423, 516
Localization, 219

M
Markov chains, 167–169, 184–185, 188, 212, 360, 443, 446–449, 453, 482, 486, 495
Markov kernels, 39, 40, 167–169, 184–185, 445–447, 453, 464, 466, 482, 533, 538
Markov processes
  Chung, 516, 538–540
  Feller, 513–515
  Hunt, 505–518, 538–540
  Ito diffusions, 455–474, 478, 497, 502, 505, 515, 525, 539
  Ito processes, 497, 515, 518, 539
  Jump-diffusions, 473–497, 505, 506, 515, 516
  piecewise deterministic, 474, 494
  step, 485–493, 496–497, 504–506, 509, 516, 517
Markov property, 168, 185, 306, 352–354, 360, 383–387, 443–454, 458–459, 501–503, 513
  for Levy processes, 342–344
Markov renewal chain, 483
Markov system, 498–505, 512–514
Martingales
  convergence, 199–213
  indistinguishability, 241
  modifications, 234–241
  stopped, 193–194, 218, 223

uniformly integrable, 183–184, 217–219, 417, 462, 519
upcrossings, 195–198
Martingales in L^p
L^p-bounded, 189, 211
square integrable, 189
Measure
absolute continuity, 31, 34, 35, 208, 370–371
atoms of, 17, 18
counting, 14, 18, 47, 91, 243, 246, 248–251, 256, 257, 272, 294, 296, 358
diffuse, 17, 18, 36, 271, 279
Dirac, 14, 17
discrete, 14, 17, 18, 21
finite, -finite, -finite, 16, 18, 35
images, 29–30, 46,51, 114
Lebesgue, 14, 17, 18, 21, 30, 32–36, 56, 65
product, 42–44, 52, 55, 85, 316
purely atomic, 17, 21, 34
restrictions, 18
singularity, 35
sigma-finite, 16, 18, 35
traces, 18, 260, 269, 278
Measurable spaces
discrete, 12, 14, 21
products of, 4–5
standard, 11, 17, 30, 33, 47, 164–166
traces, 5, 18
Modifications, 234–242, 436–437
Modified Bessel function, 367, 389
Modulus of continuity, 426, 430
Monotone class theorem, 3, 10, 13, 80–81, 160, 227
Monotone convergence theorem, 20, 22–29, 143

N

Negligibility, 17, 42, 50
Normality, 499–503

O

Occupation times, 412, 419, 424
Ornstein-Uhlenbeck process, 259, 389, 457, 470

P

Past/strict past, 174–177, 180, 235–237, 241
Perfect sets, 408
Permutations, 86, 87, 207
Point processes
compensators, 308–310
Hawkes process, 311
intensity, 477
self-exciting, 298–312
shot processes, 305–308, 310
Poisson integrals, 298–312
Poisson processes
non-stationary, 230–231, 309
characterizations for, 228, 291–293, 352
Poisson random measures
conditionally, 262, 311, 466
decompositions, 260, 272
existence, 254–256
homogeneous, 250–251, 259, 352
Laplace functional for, 244–245, 251–254, 262, 264, 286
marking, 246, 272, 535
superposition's, 260, 272, 290
traces, 260, 269, 278, 323
transformations, 249, 263–277, 535
Polish spaces, 11, 263
Positive definiteness, 68, 316, 355
Potentials, 212, 519
measures, 369–370, 376
Newtonian, 523
operators, 463, 484, 486, 518–523

Predictability
  for processes, 192, 206
  for stopping times, 206, 242, 506, 508
Predictable times, 242, 506–507
Probability spaces, 49–92, 113, 154, 160–167
  completeness, 50, 51
Product spaces, 4, 5, 12, 37–47, 52, 53, 76, 153–155, 164
Progressive processes, 176
Prohorov's theorem, 114–115
Pure-jump processes, 317–320, 330, 333, 361, 365, 368, 485

Q

Quadratic variation, 190, 198, 426–429, 527–528
Quantile functions, 56–57, 110–113
Quasi-left-continuity, 505–510, 512, 514–515, 539

R

Radial Brownian motion, 225, 379, 386–389, 399, 404–407
Radon-Nikodym theorem, 31–32, 142, 143, 145, 208–210
Random fields, 170, 289
Random measures
  additive, 277–290, 298, 536
  atoms of, 246, 263
  compound Poisson, 273–274
  Dirichlet, 288
  gamma, 282, 288
  intensities of, 261–262
Random time changes, 360, 382, 422–423, 472, 516–518, 531, 540
Random transformations, 264, 271, 448
Random variables
  distribution of, 51–56
  independence, 52–53, 82–92, 159–161
  integrable, 58, 70–72
  joint distributions, 52, 60
Random vector fields, 290
Random vectors, 67–69, 338
Random walks, 88, 327, 448, 537
Reflected Brownian motion, 386, 399, 422
Resolvent, 463–466, 483–486, 521, 539
  equation, 465, 518
Riesz decomposition, 213

S

Self-similar processes, 329
Shift operators, 340–341, 383, 498
Shot noise, 253, 258–259, 306
Sigma-algebras
  Borel, discrete, trivial, 2, 4, 7, 13, 38, 47
  generated, 2–5, 11, 17, 44, 45, 53, 75–78
  separable, 208
Singularity of measures, 35, 210
Skew Brownian motion, 422
Spherical coordinates, 328–329, 367, 387
Stable processes
  increasing, 281, 330, 338, 345, 362, 363, 368, 369, 374, 410, 476, 523
  isotropic, 332, 335, 362, 363, 368
  symmetric, 331–332, 335, 337–339
Stable random measures, 281–282
Stationarity of increments, 279, 314, 377, 536
Stieltjes processes, 525–527, 532
Stochastic base, 340, 348, 383, 455, 498–499
Stochastic differentials, 456, 528, 529
Stochastic integration, 191, 455, 525–532, 538
Stopping times
  approximation, 178, 238
  classification, 242, 507–508, 540
  foretold, 178, 237, 238

hitting times, 238, 371, 389–399, 540
predictable, 242, 297, 506–508, 540
sigma-predictable, 242, 508, 540
totally unpredictable, 242, 297, 506–508, 510, 540
Strong law of large numbers, 118–121, 123, 127, 207–208
Strong Markov property
for Brownian motion, 383–384, 388
for Markov processes, 449–451, 475, 505, 511–512, 536
for Levy processes, 343–346
for Poisson processes, 230, 296
Subdivision, 37, 233, 426–428, 525
Subordination
of stable to gamma, 362–364
of Wiener to gamma, 362, 367
of Wiener to stable, 362
Subordinators, 279, 360, 362–364, 536
Subsequences, 94–96, 102, 106, 109, 114, 115
Supermedian property, 520–524
Systems, p-system, d-system, 2–5, 10, 13, 16, 19, 51, 52, 60, 80, 146, 160

T

Tail sigma-algebra, 85–87, 207, 347–348
Time changes, 32–33, 187, 231, 261, 360, 377, 382, 399, 422, 472, 515–517, 531–532, 540
Time-homogeneity, 170, 341, 385, 446, 454
Times foretold, 178, 180, 237–238
Transition function/semigroup, 170, 268, 385–386, 445–454, 458, 479–484, 486, 488, 495, 498–500, 504, 515, 518, 538–540

U

Uniform distribution on spheres, 335, 338, 363, 367

V

Variation
cross, 458, 527–528, 530
locally finite, 219, 220
quadratic, 190, 198, 426–429, 457, 527–528
total, 37, 317, 319, 320, 330, 331, 379, 426–428, 525, 528, 532
Volatility, 224, 380

W

Wiener measure/space, 388
Wiener process
d-dimensional, 314, 322, 350, 355, 384
Levy's characterization, 293–294
symmetry, scaling, inversion, 381–382

Z

Zero-one laws
Blumenthal, 347, 354, 383, 390, 502
Hewitt-Savage, 85–89
Kolmogorov, 85–86, 124, 207, 208, 347